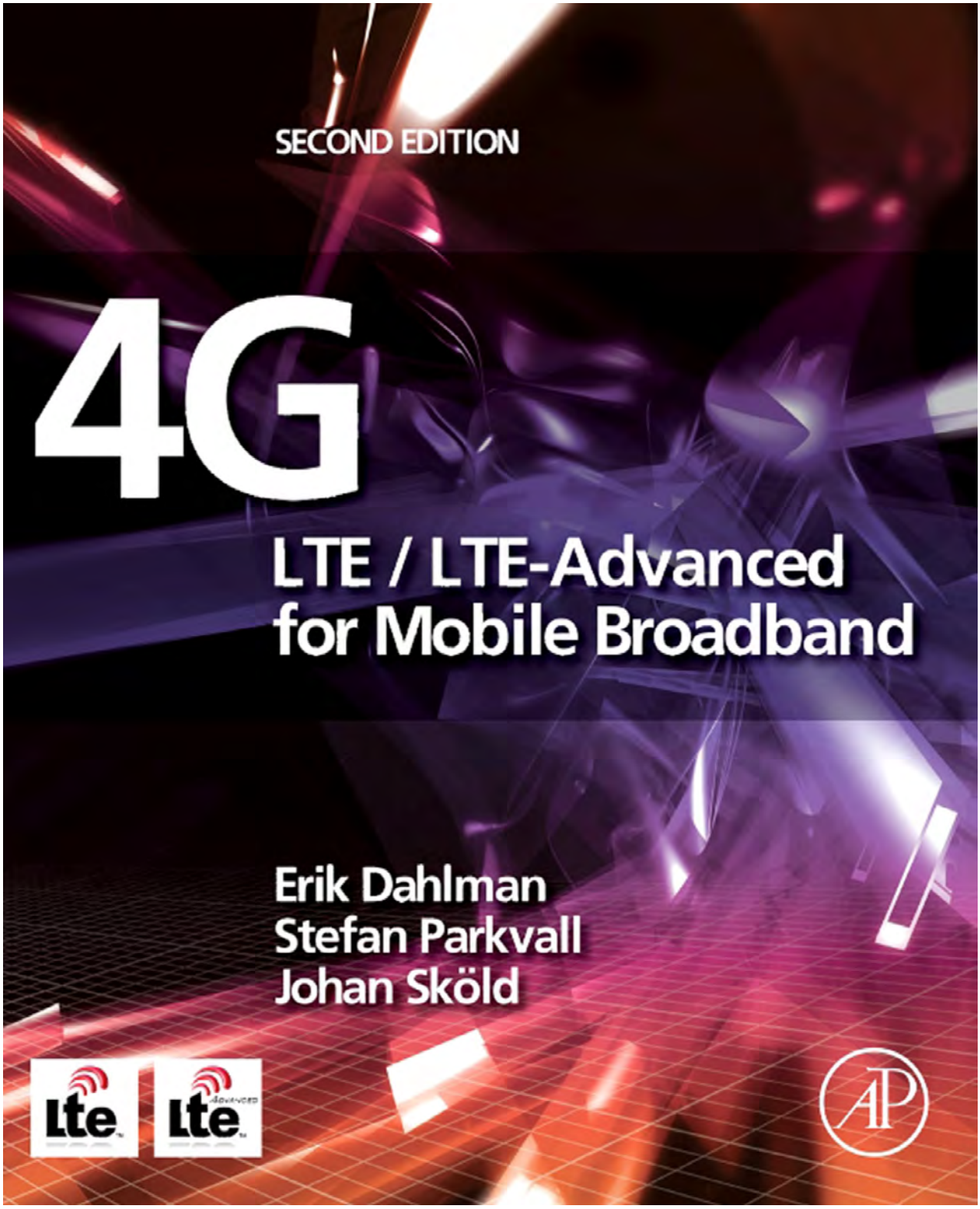


Samsung Exhibit 1017, Page 1 of 538
Samsung Electronics Co., Ltd. v. Asus Technology Licensing Inc.
IPR2024-01217

4G: LTE/LTE-Advanced for Mobile Broadband

Samsung Exhibit 1017, Page 2 of 538
Samsung Electronics Co., Ltd. v. Asus Technology Licensing Inc.
IPR2024-01217

Copyright Elsevier 2024

This page intentionally left blank

Copyright Elsevier 2024

4G: LTE/LTE-Advanced for Mobile Broadband

Second Edition

Erik Dahlman

Stefan Parkvall

Johan Sköld



AMSTERDAM • BOSTON • HEIDELBERG • LONDON
NEW YORK • OXFORD • PARIS • SAN DIEGO
SAN FRANCISCO • SINGAPORE • SYDNEY • TOKYO
Academic Press is an imprint of Elsevier



Copyright Elsevier 2014

Academic Press is an imprint of Elsevier
The Boulevard, Langford Lane, Kidlington, Oxford, OX5 1GB
225 Wyman Street, Waltham, MA 02451, USA

First published 2011
Second edition 2014

Copyright © 2014, 2011 Erik Dahlman, Stefan Parkvall and Johan Sköld. Published by Elsevier Ltd. All rights reserved.

The rights of Erik Dahlman, Stefan Parkvall and Johan Sköld to be identified as the authors of this work has been asserted in accordance with the Copyright, Designs and Patents Act 1988.

No part of this publication may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopying, recording, or any information storage and retrieval system, without permission in writing from the publisher. Details on how to seek permission, further information about the Publisher's permissions policies and our arrangement with organizations such as the Copyright Clearance Center and the Copyright Licensing Agency, can be found at our website: www.elsevier.com/permissions

This book and the individual contributions contained in it are protected under copyright by the Publisher (other than as may be noted herein).

Notices
Knowledge and best practice in this field are constantly changing. As new research and experience broaden our understanding, changes in research methods, professional practices, or medical treatment may become necessary.

Practitioners and researchers must always rely on their own experience and knowledge in evaluating and using any information, methods, compounds, or experiments described herein. In using such information or methods they should be mindful of their own safety and the safety of others, including parties for whom they have a professional responsibility.

To the fullest extent of the law, neither the Publisher nor the authors, contributors, or editors, assume any liability for any injury and/or damage to persons or property as a matter of products liability, negligence or otherwise, or from any use or operation of any methods, products, instructions, or ideas contained in the material herein.

British Library Cataloguing in Publication Data
A catalogue record for this book is available from the British Library

Library of Congress Cataloging-in-Publication Data
A catalog record for this book is available from the Library of Congress

ISBN: 978-0-12-419985-9

For information on all Academic Press publications visit our website at store.elsevier.com

Printed and bound in the United Kingdom

14 15 16 17 10 9 8 7 6 5 4 3 2 1



Copyright Elsevier 2024

Contents

Preface xiii
Acknowledgments xv
Abbreviations and Acronyms xvii

CHAPTER 1 Background of LTE 1
1.1 Introduction 1
1.2 Evolution of mobile systems before LTE 2
1.2.1 The first 3G standardization 3
1.3 ITU activities 4
1.3.1 IMT-2000 and IMT-Advanced 4
1.3.2 Spectrum for IMT systems 7
1.4 Drivers for LTE and LTE-Advanced 8
1.5 Standardization of LTE 10
1.5.1 The standardization process 10
1.5.2 The 3GPP process 12
1.5.3 The 3G evolution to 4G 14

CHAPTER 2 High Data Rates in Mobile Communication 17
2.1 High data rates: fundamental constraints 17
2.1.1 High data rates in noise-limited scenarios 19
2.1.2 Higher data rates in interference-limited scenarios 20
2.2 Higher data rates within a limited bandwidth: higher-order modulation 21
2.2.1 Higher-order modulation in combination with channel coding 22
2.2.2 Variations in instantaneous transmit power 23
2.3 Wider bandwidth including multi-carrier transmission 24
2.3.1 Multi-carrier transmission 26

CHAPTER 3 OFDM Transmission 29
3.1 Basic principles of OFDM 29
3.2 OFDM demodulation 31
3.3 OFDM implementation using IFFT/FFT processing 33
3.4 Cyclic-prefix insertion 34
3.5 Frequency-domain model of OFDM transmission 37
3.6 Channel estimation and reference symbols 37
3.7 Frequency diversity with OFDM: importance of channel coding 39
3.8 Selection of basic OFDM parameters 40
3.8.1 OFDM subcarrier spacing 41
3.8.2 Number of subcarriers 42
3.8.3 Cyclic-prefix length 43
3.9 Variations in instantaneous transmission power 43

Copyright Elsevier 2024

Samsung Exhibit 1017, Page 7 of 538
Samsung Electronics Co., Ltd. v. Asus Technology Licensing Inc.
IPR2024-01217

3.10 OFDM as a user-multiplexing and multiple-access scheme 44
3.11 Multi-cell broadcast/multicast transmission and OFDM 46

CHAPTER 4 Wider-Band "Single-Carrier" Transmission 49
4.1 Equalization against radio-channel frequency selectivity 49
4.1.1 Time-domain linear equalization 49
4.1.2 Frequency-domain equalization 52
4.1.3 Other equalizer strategies 54
4.2 Uplink FDMA with flexible bandwidth assignment 55
4.3 DFT-spread OFDM 56
4.3.1 Basic principles 57
4.3.2 DFTS-OFDM receiver 59
4.3.3 User multiplexing with DFTS-OFDM 60
4.3.4 Distributed DFTS-OFDM 61

CHAPTER 5 Multi-Antenna Techniques 63
5.1 Multi-antenna configurations 63
5.2 Benefits of multi-antenna techniques 64
5.3 Multiple receive antennas 65
5.4 Multiple transmit antennas 70
5.4.1 Transmit-antenna diversity 70
5.4.2 Means of space–time coding 72
5.4.3 Transmitter-side beam-forming 73
5.5 Spatial multiplexing 77
5.5.1 Basic principles 77
5.5.2 Precoder-based spatial multiplexing 80
5.5.3 Nonlinear receiver processing 81

CHAPTER 6 Scheduling, Link Adaptation, and Hybrid ARQ 85
6.1 Link adaptation: power and rate control 85
6.2 Channel-dependent scheduling 87
6.2.1 Downlink scheduling 87
6.2.2 Uplink scheduling 92
6.2.3 Link adaptation and channel-dependent scheduling in the frequency domain 94
6.2.4 Acquiring channel-state information 94
6.2.5 Traffic behavior and scheduling 95
6.3 Advanced retransmission schemes 96
6.4 Hybrid ARQ with soft combining 98

Copyright Elsevier 2024

CHAPTER 7 LTE Radio Access: An Overview 103
7.1 Basic principles 104
7.1.1 Transmission scheme 104
7.1.2 Channel-dependent scheduling and rate adaptation 106
7.1.3 Inter-cell interference coordination 107
7.1.4 Hybrid ARQ with soft combining 108
7.1.5 Multi-antenna transmission 108
7.1.6 Spectrum flexibility 109
7.2 LTE release 9 111
7.2.1 Multicast and broadcast support 111
7.2.2 Positioning 112
7.2.3 Dual-layer beam-forming 112
7.3 LTE release 10 and IMT-advanced 112
7.3.1 Carrier aggregation 113
7.3.2 Extended multi-antenna transmission 114
7.3.3 Relaying 114
7.3.4 Heterogeneous deployments 114
7.4 Release 11 115
7.4.1 Multi-point coordination and transmission 116
7.4.2 Enhanced control channel structure 116
7.4.3 Carrier-aggregation enhancements 117
7.4.4 Advanced receivers 117
7.5 Terminal capabilities 117

CHAPTER 8 Radio-Interface Architecture 121
8.1 Overall system architecture 121
8.1.1 Core network 121
8.1.2 Radio-access network 123
8.2 Radio protocol architecture 124
8.2.1 Radio-link control 127
8.2.2 Medium-access control 128
8.2.3 Physical layer 136
8.3 Control-plane protocols 137
8.3.1 State machine 138

CHAPTER 9 Physical Transmission Resources 141
9.1 Overall time–frequency structure 141
9.2 Normal subframes and MBSFN subframes 145
9.3 Antenna ports 147
9.3.1 Quasi-co-located antenna ports 148
9.4 Duplex schemes 148
9.4.1 Frequency-division duplex (FDD) 149

Copyright Elsevier 2024

9.4.2 Time-division duplex (TDD) 150
9.4.3 LTE and TD-SCDMA coexistence 153
9.5 Carrier aggregation 154
9.6 Frequency-domain location of LTE carriers 158

CHAPTER 10 Downlink Physical-Layer Processing 161
10.1 Transport-channel processing 161
10.1.1 Processing steps 162
10.1.2 Localized and distributed resource mapping 167
10.2 Downlink reference signals 171
10.2.1 Cell-specific reference signals 172
10.2.2 Demodulation reference signals 175
10.2.3 CSI reference signals 179
10.2.4 Quasi-co-location relations 183
10.3 Multi-antenna transmission 184
10.3.1 Transmission modes 184
10.3.2 Transmit diversity 186
10.3.3 Codebook-based precoding 189
10.3.4 Non-codebook-based precoding 193
10.3.5 Downlink multi-user MIMO 195
10.4 Downlink L1/L2 control signaling 198
10.4.1 Physical control format indicator channel 200
10.4.2 Physical hybrid-ARQ indicator channel 203
10.4.3 Physical downlink control channel 207
10.4.4 Enhanced physical downlink control channel 213
10.4.5 Blind decoding of PDCCHs and EPDCCHs 218
10.4.6 Downlink scheduling assignments 223
10.4.7 Uplink scheduling grants 233
10.4.8 Carrier aggregation and cross-carrier scheduling 237
10.4.9 Power-control commands 240

CHAPTER 11 Uplink Physical-Layer Processing 241
11.1 Transport-channel processing 241
11.1.1 Processing steps 241
11.1.2 Mapping to the physical resource 244
11.1.3 PUSCH frequency hopping 245
11.2 Uplink reference signals 248
11.2.1 Demodulation reference signals 249
11.2.2 Sounding reference signals 257
11.3 Uplink multi-antenna transmission 261
11.3.1 Precoder-based multi-antenna transmission for PUSCH 261

Copyright Elsevier 2024

11.3.2 Uplink multi-user MIMO 265
11.3.3 PUCCH transmit diversity 266
11.4 Uplink L1/L2 control signaling 267
11.4.1 Basic PUCCH structure 268
11.4.2 Uplink control signaling on PUCCH 277
11.4.3 Uplink L1/L2 control signaling on PUSCH 285
11.5 Uplink power control 288
11.5.1 Uplink power control: Some basic rules 289
11.5.2 Power control for PUCCH 291
11.5.3 Power control for PUSCH 293
11.6 Uplink timing alignment 295

CHAPTER 12 Retransmission Protocols 299
12.1 Hybrid ARQ with soft combining 300
12.1.1 Downlink hybrid ARQ 303
12.1.2 Uplink hybrid ARQ 304
12.1.3 Hybrid-ARQ timing 307
12.2 Radio-link control 314
12.2.1 Segmentation and concatenation, and reassembly of RLC SDUs 315
12.2.2 RLC retransmission 315
12.2.3 In-sequence delivery 316
12.2.4 RLC operation 317

CHAPTER 13 Scheduling and Rate Adaptation 321
13.1 Downlink scheduling 322
13.2 Uplink scheduling 324
13.2.1 Uplink priority handling 325
13.2.2 Scheduling requests 325
13.2.3 Buffer status reports 327
13.2.4 Power headroom reports 328
13.3 Timing of scheduling assignments/grants 329
13.3.1 Downlink scheduling timing 330
13.3.2 Uplink scheduling timing 331
13.4 Semi-persistent scheduling 332
13.5 Scheduling for half-duplex FDD 333
13.6 Discontinuous reception (DRX) and component carrier deactivation 335
13.7 Channel-state information 337
13.7.1 Channel-quality indication 338
13.7.2 Rank indicator and precoder matrix indication 338
13.7.3 Periodic and aperiodic CSI reports 340
13.7.4 CSI estimation 344

Copyright Elsevier 2024

CHAPTER 14 Access Procedures 347
14.1 Acquisition and cell search 347
14.1.1 Overview of LTE cell search 347
14.1.2 PSS structure 349
14.1.3 SSS structure 350
14.2 System information 351
14.2.1 Master information block and BCH transmission 351
14.2.2 System-information blocks 354
14.3 Random access 358
14.3.1 Step 1: Random-access preamble transmission 360
14.3.2 Step 2: Random-access response 366
14.3.3 Step 3: Terminal identification 367
14.3.4 Step 4: Contention resolution 368
14.4 Paging 369

CHAPTER 15 Multi-Point Coordination and Transmission 371
15.1 Release-8 inter-cell interference coordination 374
15.2 Releases-10/11 multi-point coordination/transmission 375
15.2.1 Multi-point coordination 376
15.2.2 Multi-point transmission 381
15.2.3 Uplink multi-point coordination/reception 385

CHAPTER 16 Heterogeneous Deployments 387
16.1 Interference scenarios in heterogeneous deployments 388
16.2 Heterogeneous deployments using rel-8 functionality 392
16.3 Frequency-domain partitioning 393
16.4 Time-domain partitioning 394
16.5 Shared cell 398
16.6 Closed subscriber groups 401

CHAPTER 17 Multimedia Broadcast/Multicast Services 403
17.1 Architecture 404
17.2 Overall channel structure and physical-layer processing 406
17.3 Scheduling of MBMS services 408

CHAPTER 18 Relaying 413
18.1 Relays in LTE 414
18.2 Overall architecture 415
18.3 Backhaul design for inband relaying 417
18.3.1 Access-link hybrid-ARQ operation 418
18.3.2 Backhaul-link hybrid-ARQ operation 419

Copyright Elsevier 2024

18.3.3 Backhaul downlink control signaling ... 421
18.3.4 Reference signals for the backhaul link ... 424
18.3.5 Backhaul–access link timing ... 425

CHAPTER 19 Spectrum and RF Characteristics ... 431
19.1 Spectrum for LTE ... 431
19.1.1 Spectrum defined for IMT systems by the ITU-R ... 432
19.1.2 Frequency bands for LTE ... 432
19.1.3 New frequency bands ... 437
19.2 Flexible spectrum use ... 438
19.3 Flexible channel bandwidth operation ... 439
19.4 Carrier aggregation for LTE ... 441
19.5 Operation in non-contiguous spectrum ... 445
19.6 Multi-standard radio base stations ... 447
19.7 Overview of RF requirements for LTE ... 451
19.7.1 Transmitter characteristics ... 452
19.7.2 Receiver characteristics ... 453
19.7.3 Regional requirements ... 453
19.7.4 Band-specific terminal requirements through network signaling ... 454
19.7.5 Base-station classes ... 455
19.8 Output power level requirements ... 456
19.8.1 Base-station output power and dynamic range ... 456
19.8.2 Terminal output power and dynamic range ... 456
19.9 Transmitted signal quality ... 457
19.9.1 EVM and frequency error ... 457
19.9.2 Terminal in-band emissions ... 458
19.9.3 Base-station time alignment ... 458
19.10 Unwanted emissions requirements ... 458
19.10.1 Implementation aspects ... 459
19.10.2 Spectrum emission mask ... 459
19.10.3 Adjacent channel leakage ratio ... 461
19.10.4 Spurious emissions ... 463
19.10.5 Occupied bandwidth ... 464
19.10.6 Transmitter intermodulation ... 464
19.11 Sensitivity and dynamic range ... 464
19.12 Receiver susceptibility to interfering signals ... 465
19.13 Multi-band-capable base stations ... 467
19.14 RF requirements for relays ... 470

CHAPTER 20 Performance ... 473
20.1 Performance assessment ... 473
20.1.1 End-user perspective of performance ... 474
20.1.2 Operator perspective ... 475

Copyright Elsevier 2024

20.2 Performance in terms of peak data rates and latency 476
20.3 Performance evaluation of LTE-advanced 477
20.3.1 Models and assumptions 477
20.3.2 Evaluation criteria 480
20.3.3 Performance numbers 481
20.4 Conclusion 484

CHAPTER 21 Future Radio Access—Final Thoughts 485
21.1 Continued LTE evolution beyond release 11 486
21.1.1 Further enhanced local-area access 487
21.1.2 Enhanced multi-antenna/multi-point enhancements 490
21.1.3 Machine-type communication 491
21.1.4 Device-to-device communication 493
21.2 New radio access 494
21.2.1 New frequency ranges 496
21.2.2 New ways of assigning spectrum 496
21.2.3 Massive antenna configurations 497
21.2.4 Ultra-dense deployments 498
21.3 Final thoughts 498

References 499
Index 505

Copyright Elsevier 2024

Preface

During the past years, there has been a quickly rising interest in radio access technologies for providing mobile as well as nomadic and fixed services for voice, video, and data. This proves that the difference in design, implementation, and use between telecom and datacom technologies is also becoming more blurred. What used to be a mobile phone used for voice communication is today increasingly becoming the main data communication device for end-users, providing web browsing, social networking, and many other services.

Today, the most widespread radio access technology for mobile communication is digital cellular, with the number of subscribers exceeding six billion in 2013, quickly approaching an average of one subscription for every person on Earth. It has emerged from early deployments of an expensive voice service for a few car-borne users to today's widespread use of mobile-communication devices that provide a range of mobile services and often include camera, MP3 player, and PDA functions. With this widespread use and increasing interest in mobile communication, there is a continuing evolution ongoing.

This book describes LTE, developed in 3GPP (*Third Generation Partnership Project*) and providing true 4G broadband mobile access, starting from the first version in release 8 and through the continuing evolution to release 11, the latest version of LTE. Releases 10 and 11, also known as LTE-Advanced, is of particular interest as it is the major technology approved by the ITU as fulfilling the IMT-Advanced requirements. The description in this book is based on LTE release 11 and thus provides a complete description of the LTE-Advanced radio access from the bottom up.

Chapter 1 gives the background to LTE and its evolution, looking also at the different standards bodies and organizations involved in the process of defining 4G. It also gives a discussion of the reasons and driving forces behind the evolution.

Chapters 2–6 provide a deeper insight into some of the technologies that are part of LTE and its evolution. Because of its generic nature, these chapters can be used as a background not only for LTE as described in this book, but also for readers who want to understand the technology behind other systems, such as WCDMA/HSPA, WiMAX, and CDMA2000.

Chapters 7–19 constitute the main part of the book. As a start, a technical overview of LTE is given, where the most important technology components are introduced based on the generic technologies described in previous chapters. The following chapters provide a detailed description of the protocol structure, the downlink and uplink transmission schemes, and the associated mechanisms for scheduling, retransmission, and interference handling. Broadcast operation and relaying are also described. This is followed by a discussion of the spectrum flexibility and the associated requirements from an RF perspective.

In Chapter 20, an assessment is made on LTE, first by looking at the performance of LTE-Advanced from a system point of view.

Finally, looking into the future in Chapter 21, it will be seen that the evolution does not stop with LTE-Advanced but that new features are continuously added to LTE in order to meet future requirements.

Copyright Elsevier 2024

Samsung Exhibit 1017, Page 15 of 538
Samsung Electronics Co., Ltd. v. Asus Technology Licensing Inc.
IPR2024-01217

This page intentionally left blank

Copyright Elsevier 202

Acknowledgments

We thank all our colleagues at Ericsson for assisting in this project by helping with contributions to the book, giving suggestions and comments on the contents, and taking part in the huge team effort of developing LTE.

The standardization process involves people from all parts of the world, and we acknowledge the efforts of our colleagues in the wireless industry in general and in 3GPP RAN in particular. Without their work and contributions to the standardization, this book would not have been possible.

Finally, we are immensely grateful to our families for bearing with us and supporting us during the long process of writing this book.

Copyright Elsevier 2024

Samsung Exhibit 1017, Page 17 of 538
Samsung Electronics Co., Ltd. v. Asus Technology Licensing Inc.
IPR2024-01217

This page intentionally left blank

Copyright Elsevier 202

Abbreviations and Acronyms

3GPP	Third Generation Partnership Project
ACIR	Adjacent Channel Interference Ratio
ACK	Acknowledgement (in ARQ protocols)
ACLR	Adjacent Channel Leakage Ratio
ACS	Adjacent Channel Selectivity
AM	Acknowledged Mode (RLC configuration)
AMC	Adaptive Modulation and Coding
A-MPR	Additional Maximum Power Reduction
AMPS	Advanced Mobile Phone System
AQPSK	Adaptive QPSK
ARI	Acknowledgement Resource Indicator
ARIB	Association of Radio Industries and Businesses
ARQ	Automatic Repeat-reQuest
AS	Access Stratum
ATIS	Alliance for Telecommunications Industry Solutions
AWGN	Additive White Gaussian Noise
BC	Band Category
BCCH	Broadcast Control Channel
BCH	Broadcast Channel
BER	Bit-Error Rate
BLER	Block-Error Rate
BM-SC	Broadcast Multicast Service Center
BPSK	Binary Phase-Shift Keying
BS	Base Station
BSC	Base Station Controller
BTS	Base Transceiver Station
CA	Carrier Aggregation
CACLR	Cumulative Adjacent Channel Leakage Ratio
CC	Convolutional Code (in the context of coding), or Component Carrier (in the context of carrier aggregation).
CCCH	Common Control Channel
CCE	Control Channel Element
CCSA	China Communications Standards Association
CDD	Cyclic-Delay Diversity
CDF	Cumulative Density Function
CDM	Code-Division Multiplexing
CDMA	Code-Division Multiple Access
CEPT	European Conference of Postal and Telecommunications Administrations
CN	Core Network

Copyright Elsevier 2024

Samsung Exhibit 1017, Page 19 of 538
Samsung Electronics Co., Ltd. v. Asus Technology Licensing Inc.
IPR2024-01217

CoMP	Coordinated Multi-Point Transmission/Reception
CP	Cyclic Prefix
CPC	Continuous Packet Connectivity
CQI	Channel-Quality Indicator
C-RAN	Centralized RAN
CRC	Cyclic Redundancy Check
C-RNTI	Cell Radio-Network Temporary Identifier
CRS	Cell-specific Reference Signal
CS	Circuit Switched (or Cyclic Shift)
CS	Capability Set (for MSR base stations)
CSA	Common Subframe Allocation
CSG	Closed Subscriber Group
CSI	Channel-State Information
CSI-IM	CSI Interference Measurement
CSI-RS	CSI Reference Signals
CW	Continuous Wave
DAI	Downlink Assignment Index
DCCH	Dedicated Control Channel
DCH	Dedicated Channel
DCI	Downlink Control Information
DFE	Decision-Feedback Equalization
DFT	Discrete Fourier Transform
DFTS-OFDM	DFT-Spread OFDM (DFT-precoded OFDM, see also SC-FDMA)
DL	Downlink
DL-SCH	Downlink Shared Channel
DM-RS	Demodulation Reference Signal
DRX	Discontinuous Reception
DTCH	Dedicated Traffic Channel
DTX	Discontinuous Transmission
DwPTS	The Downlink Part of the Special Subframe (for TDD operation).
ECCE	Enhanced Control Channel Element
EDGE	Enhanced Data Rates for GSM Evolution, Enhanced Data Rates for Global Evolution
EGPRS	Enhanced GPRS
eNB	eNodeB
eNodeB	E-UTRAN NodeB
EPC	Evolved Packet Core
EPDCCH	Enhanced Physical Downlink Control Channel
EPS	Evolved Packet System
EREG	Enhanced Resource-Element Group
ETSI	European Telecommunications Standards Institute
E-UTRA	Evolved UTRA
E-UTRAN	Evolved UTRAN

Copyright Elsevier 2024

EV-DO	Evolution-Data Only (of CDMA2000 1x)
EV-DV	Evolution-Data and Voice (of CDMA2000 1x)
EVM	Error Vector Magnitude
FACH	Forward Access Channel
FCC	Federal Communications Commission
FDD	Frequency Division Duplex
FDM	Frequency-Division Multiplex
FDMA	Frequency-Division Multiple Access
FEC	Forward Error Correction
FeICIC	Further Enhanced Inter—Cell Interference Coordiantion
FFT	Fast Fourier Transform
FIR	Finite Impulse Response
FPLMTS	Future Public Land Mobile Telecommunications Systems
FRAMES	Future Radio Wideband Multiple Access Systems
FSTD	Frequency Switched Transmit Diversity
GERAN	GSM/EDGE Radio Access Network
GGSN	Gateway GPRS Support Node
GP	Guard Period (for TDD operation)
GPRS	General Packet Radio Services
GPS	Global Positioning System
GSM	Global System for Mobile Communications
HARQ	Hybrid ARQ
HII	High-Interference Indicator
HLR	Home Location Register
HRPD	High Rate Packet Data
HSPA	High-Speed Packet Access
HSS	Home Subscriber Server
HS-SCCH	High-Speed Shared Control Channel
ICIC	Inter-Cell Interference Coordination
ICS	In-Channel Selectivity
ICT	Information and Communication Technologies
IDFT	Inverse DFT
IEEE	Institute of Electrical and Electronics Engineers
IFDMA	Interleaved FDMA
IFFT	Inverse Fast Fourier Transform
IMT-2000	International Mobile Telecommunications 2000 (ITU's name for the family of 3G standards).
IMT-Advanced	International Mobile Telecommunications Advanced (ITU's name for the family of 4G standards).
IP	Internet Protocol
IR	Incremental Redundancy
IRC	Interference Rejection Combining
ITU	International Telecommunications Union

Copyright Elsevier 2024

ITU-R	International Telecommunications Union-Radiocommunications Sector
J-TACS	Japanese Total Access Communication System
LAN	Local Area Network
LCID	Logical Channel Index
LDPC	Low-Density Parity Check Code
LSA	Licensed Shared Access
LTE	Long-Term Evolution
MAC	Medium Access Control
MAN	Metropolitan Area Network
MBMS	Multimedia Broadcast/Multicast Service
MBMS-GW	MBMS Gateway
MB-MSR	Multi-Band Multi Standard Radio (base station)
MBS	Multicast and Broadcast Service
MBSFN	Multicast-Broadcast Single Frequency Network
MC	Multi-Carrier
MCCH	MBMS Control Channel
MCE	MBMS Coordination Entity
MCH	Multicast Channel
MCS	Modulation and Coding Scheme
MDHO	Macro-Diversity Handover
MIB	Master Information Block
MIMO	Multiple-Input Multiple-Output
ML	Maximum Likelihood
MLSE	Maximum-Likelihood Sequence Estimation
MME	Mobility Management Entity
MMS	Multimedia Messaging Service
MMSE	Minimum Mean Square Error
MPR	Maximum Power Reduction
MRC	Maximum Ratio Combining
MSA	MCH Subframe Allocation
MSC	Mobile Switching Center
MSI	MCH Scheduling Information
MSP	MCH Scheduling Period
MSR	Multi-Standard Radio
MSS	Mobile Satellite Service
MTC	Machine-Type Communication
MTCH	MBMS Traffic Channel
MU-MIMO	Multi-User MIMO
MUX	Multiplexer or Multiplexing
NAK, NACK	Negative Acknowledgement (in ARQ protocols)
NAS	Non-Access Stratum (a functional layer between the core network and the terminal that supports signaling and user data transfer).
NDI	New-Data Indicator

Copyright Elsevier 2024

NMT	Nordisk MobilTelefon (Nordic Mobile Telephony)
NodeB	NodeB, a logical node handling transmission/reception in multiple cells. Commonly, but not necessarily, corresponding to a base station.
NS	Network Signaling
NSPS	National Security and Public Safety
OCC	Orthogonal Cover Code
OFDM	Orthogonal Frequency-Division Multiplexing
OFDMA	Orthogonal Frequency-Division Multiple Access
OI	Overload Indicator
OOB	Out-Of-Band (emissions)
PAPR	Peak-to-Average Power Ratio
PAR	Peak-to-Average Ratio (same as PAPR)
PARC	Per-Antenna Rate Control
PBCH	Physical Broadcast Channel
PCCH	Paging Control Channel
PCFICH	Physical Control Format Indicator Channel
PCG	Project Coordination Group (in 3GPP)
PCH	Paging Channel
PCRF	Policy and Charging Rules Function
PCS	Personal Communications Systems
PDA	Personal Digital Assistant
PDC	Personal Digital Cellular
PDCCH	Physical Downlink Control Channel
PDCP	Packet Data Convergence Protocol
PDSCH	Physical Downlink Shared Channel
PDN	Packet Data Network
PDU	Protocol Data Unit
PF	Proportional Fair (a type of scheduler)
P-GW	Packet-Data Network Gateway (also PDN-GW)
PHICH	Physical Hybrid-ARQ Indicator Channel
PHS	Personal Handy-Phone System
PHY	Physical Layer
PMCH	Physical Multicast Channel
PMI	Precoding-Matrix Indicator
POTS	Plain Old Telephony Services
PRACH	Physical Random Access Channel
PRB	Physical Resource Block
P-RNTI	Paging RNTI
PS	Packet Switched
PSK	Phase Shift Keying
PSS	Primary Synchronization Signal
PSTN	Public Switched Telephone Networks
PUCCH	Physical Uplink Control Channel

Copyright Elsevier 2024

PUSC	Partially Used Subcarriers (for WiMAX)
PUSCH	Physical Uplink Shared Channel
QAM	Quadrature Amplitude Modulation
QCL	Quasi-Co-Location
QoS	Quality-of-Service
QPP	Quadrature Permutation Polynomial
QPSK	Quadrature Phase-Shift Keying
RAB	Radio Access Bearer
RACH	Random Access Channel
RAN	Radio Access Network
RA-RNTI	Random Access RNTI
RAT	Radio Access Technology
RB	Resource Block
RE	Resource Element
REG	Resource-Element Group
RF	Radio Frequency
RI	Rank Indicator
RIT	Radio Interface Technology
RLC	Radio Link Control
RNC	Radio Network Controller
RNTI	Radio-Network Temporary Identifier
RNTP	Relative Narrowband Transmit Power
ROHC	Robust Header Compression
R-PDCCH	Relay Physical Downlink Control Channel
RR	Round-Robin (a type of scheduler)
RRC	Radio Resource Control
RRM	Radio Resource Management
RS	Reference Symbol
RSPC	IMT-2000 Radio Interface Specifications
RSRP	Reference Signal Received Power
RSRQ	Reference Signal Received Quality
RTP	Real Time Protocol
RTT	Round-Trip Time
RV	Redundancy Version
RX	Receiver
S1	The interface between eNodeB and the Evolved Packet Core.
S1-c	The control-plane part of S1.
S1-u	The user-plane part of S1.
SAE	System Architecture Evolution
SCM	Spatial Channel Model
SDMA	Spatial Division Multiple Access
SDO	Standards Developing Organization
SDU	Service Data Unit
SEM	Spectrum Emissions Mask

Copyright Elsevier 2024

SF	Spreading Factor
SFBC	Space-Frequency Block Coding
SFN	Single-Frequency Network (in general, see also MBSFN) or System Frame Number (in 3GPP).
SFTD	Space–Frequency Time Diversity
SGSN	Serving GPRS Support Node
S-GW	Serving Gateway
SI	System Information Message
SIB	System Information Block
SIC	Successive Interference Combining
SIM	Subscriber Identity Module
SINR	Signal-to-Interference-and-Noise Ratio
SIR	Signal-to-Interference Ratio
SI-RNTI	System Information RNTI
SMS	Short Message Service
SNR	Signal-to-Noise Ratio
SOHO	Soft Handover
SORTD	Spatial Orthogonal-Resource Transmit Diversity
SR	Scheduling Request
SRS	Sounding Reference Signal
SSS	Secondary Synchronization Signal
STBC	Space–Time Block Coding
STC	Space–Time Coding
STTD	Space-Time Transmit Diversity
SU-MIMO	Single-User MIMO
TACS	Total Access Communication System
TCP	Transmission Control Protocol
TC-RNTI	Temporary C-RNTI
TD-CDMA	Time-Division Code-Division Multiple Access
TDD	Time-Division Duplex
TDM	Time-Division Multiplexing
TDMA	Time-Division Multiple Access
TD-SCDMA	Time-Division-Synchronous Code-Division Multiple Access
TF	Transport Format
TIA	Telecommunications Industry Association
TM	Transparent Mode (RLC configuration), Transmission Mode (physical layer)
TR	Technical Report
TS	Technical Specification
TSG	Technical Specification Group
TTA	Telecommunications Technology Association
TTC	Telecommunications Technology Committee
TTI	Transmission Time Interval
TX	Transmitter

Copyright Elsevier 2024

UCI	Uplink Control Information
UE	User Equipment, the 3GPP name for the mobile terminal.
UEM	Unwanted Emissions Mask
UL	Uplink
UL-SCH	Uplink Shared Channel
UM	Unacknowledged Mode (RLC configuration)
UMB	Ultra Mobile Broadband
UMTS	Universal Mobile Telecommunications System
UpPTS	The uplink part of the special subframe (for TDD operation).
US-TDMA	US Time-Division Multiple Access Standard
UTRA	Universal Terrestrial Radio Access
UTRAN	Universal Terrestrial Radio Access Network
VAMOS	Voice Services over Adaptive Multi-User Channels
VoIP	Voice-over-IP
VRB	Virtual Resource Block
WAN	Wide Area Network
WARC	World Administrative Radio Congress
WCDMA	Wideband Code-Division Multiple Access
WCS	Wireless Communications Service
WG	Working Group
WiMAX	Worldwide Interoperability for Microwave Access
WLAN	Wireless Local Area Network
WMAN	Wireless Metropolitan Area Network
WP5D	Working Party 5D
WRC	World Radiocommunication Conference
X2	The interface between eNodeBs.
ZC	Zadoff-Chu
ZF	Zero Forcing

Copyright Elsevier 2024

CHAPTER

Background of LTE

1

1.1 Introduction

Mobile communication has become an everyday commodity. In the last decades, it has evolved from being an expensive technology for a few selected individuals to today's ubiquitous systems used by a majority of the world's population. From the first experiments with radio communication by Guglielmo Marconi in the 1890s, the road to truly mobile radio communication has been quite long. To understand the complex mobile-communication systems of today, it is important to understand where they came from and how cellular systems have evolved. The task of developing mobile technologies has also changed, from being a national or regional concern to becoming an increasingly complex task undertaken by global standards-developing organizations such as the *Third Generation Partnership Project* (3GPP) and involving thousands of people.

Mobile communication technologies are often divided into generations (see Figure 1.1), with 1G being the analog mobile radio systems of the 1980s, 2G the first digital mobile systems, and 3G the first mobile systems handling broadband data. The next generation, 4G or *Long-Term Evolution* (LTE), provides even better support for mobile broadband. Further evolution steps of 4G LTE will be taken within the next few years. In a longer term perspective, around 2020 one may enter into what some would call "5G" radio access, as discussed in Chapter 21. This continuing race of increasing sequence numbers for mobile system generations is in fact just a matter of labels. What is important is the actual system capabilities and how they have evolved, which is the topic of this chapter.

Later releases of LTE are sometimes known as LTE-Advanced, but it is important to point out that LTE and LTE-Advanced are the same technology. The label "Advanced" was primarily added to highlight the relationship between LTE release 10 (LTE-Advanced) and ITU/IMT-Advanced, as discussed later. This does not make LTE-Advanced a different system than LTE, and it is not in any way the final evolutionary step to be taken for LTE. Another important aspect is that the developmental work on LTE and LTE-Advanced is performed as a continuing task within 3GPP, the same forum that developed the first 3G system (WCDMA/HSPA). This book covers LTE and LTE-Advanced up to and including 3GPP release 11, with an outlook to what is expected in release 12 and future radio-access in general.

4G: LTE/LTE-Advanced for Mobile Broadband. http://dx.doi.org/10.1016/B978-0-12-419985-9.00001-5
Copyright © 2014 Erik Dahlman, Stefan Parkvall and Johan Sköld. Published by Elsevier Ltd. All rights reserved.

Copyright Elsevier 2024

Samsung Exhibit 1017, Page 27 of 538
Samsung Electronics Co., Ltd. v. Asus Technology Licensing Inc.
IPR2024-01217

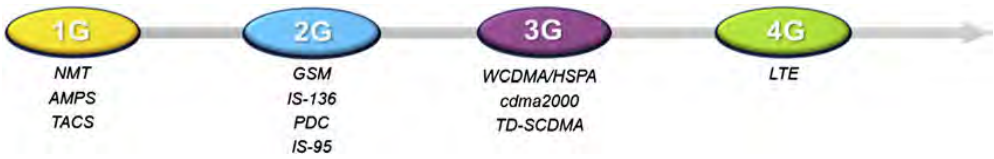


FIGURE 1.1

Generations of mobile communication systems

This chapter describes the background for the development of the LTE system in terms of events, activities, organizations and other factors that have played an important role. First, the technologies and mobile systems leading up to the starting point for 3G mobile systems will be discussed. Next, international activities in the ITU that were part of shaping 3G and the 3G evolution and the market and technology drivers behind LTE will be discussed. The final part of the chapter describes the standardization process that provided the detailed specification work leading to the LTE systems deployed and in operation today.

1.2 Evolution of mobile systems before LTE

The US *Federal Communications Commission* (FCC) approved the first commercial car-borne telephony service in 1946, operated by AT&T. In 1947 AT&T also introduced the cellular concept of reusing radio frequencies, which became fundamental to all subsequent mobile-communication systems. Similar systems were operated by several monopoly telephone administrations and wire-line operators during the 1950s and 1960s, using bulky and power-hungry equipment and providing car-borne services for a very limited number of users.

The big uptake of subscribers and usage came when mobile communication became an international concern involving several interested parties, in the beginning mainly the operators. The first international mobile communication systems were started in the early 1980s; the best-known ones are NMT that began in the Nordic countries, AMPS in North America, TACS in Europe, and J-TACS in Japan. Equipment was still bulky, mainly car-borne, and voice quality was often inconsistent, with "cross-talk" between users being a common problem. With NMT came the concept of "roaming," providing a service for users traveling outside the area of their "home" operator. This also created a larger market for mobile phones, attracting more companies into the mobile-communication business.

The analog first-generation cellular systems supported "plain old telephony services" (POTS)—that is, voice with some related supplementary services. With the advent of digital communication during the 1980s, the opportunity to develop a second generation of mobile-communication standards and systems, based on digital technology, surfaced. With digital technology came an opportunity to increase the capacity of the systems, to give a more consistent quality of the service, and to develop much more attractive and truly mobile devices.

Copyright Elsevier 2024

In Europe in the mid-1980s the GSM (originally *Groupe Spécial Mobile*, later *Global System for Mobile Communications*) project to develop a pan-European mobile-telephony system was initiated by the telecommunication administrations in CEPT[1] and later continued within the new *European Telecommunication Standards Institute* (ETSI). The GSM standard was based on *Time-Division Multiple Access* (TDMA), as were the US-TDMA standard and the Japanese PDC standard that were introduced in the same time frame. A somewhat later development of a *Code-Division Multiple Access* (CDMA) standard called IS-95 was completed in the USA in 1993.

All these standards were "narrowband" in the sense that they targeted "low-bandwidth" services such as voice. With the second-generation digital mobile communications came the opportunity to provide data services over the mobile-communication networks. The primary data services introduced in 2G were text messaging (*Short Message Services,* SMS) and circuit-switched data services enabling email and other data applications, initially at a modest peak data rate of 9.6 kbit/s. Higher data rates were introduced later in evolved 2G systems by assigning multiple time slots to a user and through modified coding schemes.

Packet data over cellular systems became a reality during the second half of the 1990s, with *General Packet Radio Services* (GPRS) introduced in GSM and packet data also added to other cellular technologies such as the Japanese PDC standard. These technologies are often referred to as 2.5G. The success of the wireless data service iMode in Japan, which included a complete "ecosystem" for service delivery, charging, etc. gave a very clear indication of the potential for applications over packet data in mobile systems, in spite of the fairly low data rates supported at the time.

With the advent of 3G and the higher-bandwidth radio interface of UTRA (*Universal Terrestrial Radio Access*) came possibilities for a range of new services that were only hinted at with 2G and 2.5G. The 3G/UTRA radio access development is today handled in 3GPP. However, the initial steps for 3G were taken in the early 1990s, long before 3GPP was formed.

What also set the stage for 3G was the internationalization of cellular standards. GSM was a pan-European project, but it quickly attracted worldwide interest when the GSM standard was deployed in a number of countries outside Europe. A global standard gains in economy of scale, since the market for products becomes larger. This has driven a much tighter international cooperation around 3G cellular technologies than for the earlier generations.

1.2.1 The first 3G standardization

Work on a third-generation mobile communication system started in ITU (*International Telecommunication Union*) in the 1980s, first under the label *Future Public Land Mobile Telecommunications Systems* (FPLMTS), later changed to IMT-2000 [1]. The World Administrative Radio Congress WARC-92 identified 230 MHz of spectrum for IMT-2000 on

[1]The European Conference of Postal and Telecommunications Administrations (CEPT) consists of the telecom administrations from 48 countries.

Copyright Elsevier 2024

a worldwide basis. Of these 230 MHz, 2 × 60 MHz was identified as paired spectrum for FDD (*Frequency-Division Duplex*) and 35 MHz as unpaired spectrum for TDD (*Time-Division Duplex*), both for terrestrial use. Some spectrum was also set aside for satellite services. With that, the stage was set to specify IMT-2000.

In parallel with the widespread deployment and evolution of 2G mobile-communication systems during the 1990s, substantial efforts were put into 3G research activities worldwide. In Europe, a number of partially EU-funded projects resulted in a multiple access concept that included a Wideband CDMA component that was input to ETSI in 1996. In Japan, the *Association of Radio Industries and Businesses* (ARIB) was at the same time defining a 3G wireless communication technology based on Wideband CDMA, and in the USA a Wideband CDMA concept called WIMS was developed within the T1.P1[2] committee. South Korea also started work on Wideband CDMA at this time.

When the standardization activities for 3G started in ETSI in 1996, there were WCDMA concepts proposed both from a European research project (FRAMES) and from the ARIB standardization in Japan. The Wideband CDMA proposals from Europe and Japan were merged and came out as part of the winning concept in early 1998 in the European work on Universal Mobile Telecommunication Services (UMTS), which was the European name for 3G. Standardization of WCDMA continued in parallel in several standards groups until the end of 1998, when the *Third Generation Partnership Project* (3GPP) was formed by standards-developing organizations from all regions of the world. This solved the problem of trying to maintain parallel development of aligned specifications in multiple regions. The present organizational partners of 3GPP are ARIB (Japan), CCSA (China), ETSI (Europe), ATIS (USA), TTA (South Korea), and TTC (Japan).

At this time, when the standardization bodies were ready to put the details into the 3GPP specifications, work on 3G mobile systems had already been ongoing for some time in the international arena within the ITU-R. That work was influenced by and also provided a broader international framework for the standardization work in 3GPP.

1.3 ITU activities

1.3.1 IMT-2000 and IMT-Advanced

ITU-R *Working Party 5D* (WP5D) has the responsibility for IMT systems, which is the umbrella name for 3G (IMT-2000) and 4G (IMT-Advanced). WP5D does not write technical specifications for IMT, but has kept the roles of defining IMT in cooperation with the regional standardization bodies and maintaining a set of recommendations for IMT-2000 and IMT-Advanced. Recently, ITU-R has initiated an activity, sometimes referred to as IMT-2020, to look at the development of IMT technologies beyond what is defined for IMT-2000 and IMT-Advanced.

[2]The T1.P1 committee was part of T1, which presently has joined the ATIS standardization organization.

Copyright Elsevier 2024

The main IMT-2000 recommendation is ITU-R M.1457 [2], which identifies the IMT-2000 *radio interface specifications* (RSPC). The recommendation contains a "family" of radio interfaces, all included on an equal basis. The family of six terrestrial radio interfaces is illustrated in Figure 1.2, which also shows the *Standards Developing Organizations* (SDO) or Partnership Projects that produce the specifications. In addition, there are several IMT-2000 satellite radio interfaces defined that are not illustrated in Figure 1.2.

For each radio interface, M.1457 contains an overview of that radio interface, followed by a list of references to the detailed specifications. The actual specifications are maintained by the individual SDOs and M.1457 provides references to the specifications transposed and maintained by each SDO.

With the continuing development of the IMT-2000 radio interfaces, including the evolution of UTRA to Evolved UTRA, the ITU recommendations also need to be updated. ITU-R WP5D continuously revises recommendation M.1457 and at the time of writing it is in its eleventh version. Input to the updates is provided by the SDOs and Partnership Projects writing the standards. In the latest revision of ITU-R M.1457, LTE (or E-UTRA) is included in the family through the 3GPP family members for UTRA FDD and TDD, as shown in Figure 1.2.

IMT-Advanced is the term used for systems that include new radio interfaces supporting the new capabilities of systems beyond IMT-2000, as demonstrated with the "van diagram" in Figure 1.3. The step into IMT-Advanced capabilities is seen by ITU-R as the step into 4G, the next generation of mobile technologies after 3G.

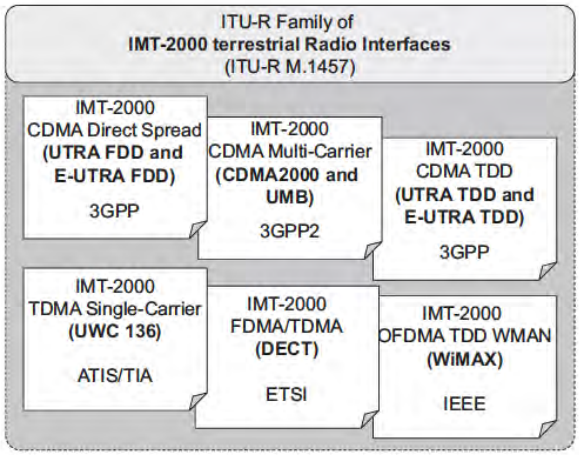


FIGURE 1.2

The definition of IMT-2000 in ITU-R

Copyright Elsevier 2024

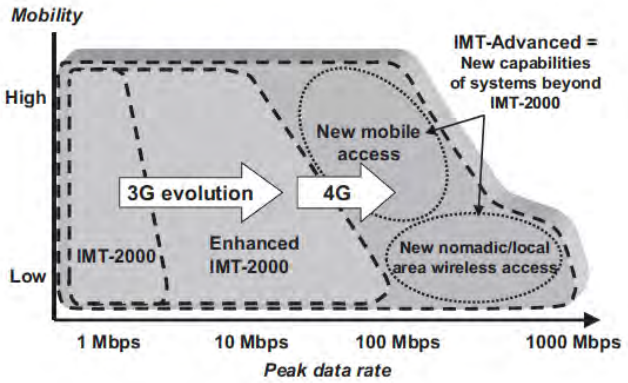


FIGURE 1.3

Illustration of capabilities of IMT-2000 and IMT-Advanced, based on the framework described in ITU-R Recommendation M.1645 [47]

The process for defining IMT-Advanced was set by ITU-R WP5D [3] and was quite similar to the process used in developing the IMT-2000 recommendations. ITU-R first concluded studies for IMT-Advanced of services and technologies, market forecasts, principles for standardization, estimation of spectrum needs, and identification of candidate frequency bands [4]. Evaluation criteria were agreed, where proposed technologies were to be evaluated according to a set of minimum technical requirements. All ITU members and other organizations were then invited to the process through a circular letter [5] in March 2008. After submission of six candidate technologies in 2009, an evaluation was performed in cooperation with external bodies such as standards-developing organizations, industry fora, and national groups.

An evolution of LTE as developed by 3GPP was submitted as one candidate to the ITU-R evaluation. While actually being a new release (release 10) of the LTE system and thus an integral part of the continuing LTE development, the candidate was named LTE-Advanced for the purpose of ITU submission. 3GPP also set up its own set of technical requirements for LTE-Advanced, with the ITU-R requirements as a basis. The specifics of LTE-Advanced will be described in more detail as part of the description of LTE later in this book. The performance evaluation of LTE-Advanced for the ITU-R submission is described further in Chapter 20.

The target of the process was always harmonization of the candidates through consensus building. ITU-R determined in October 2010 that two technologies would be included in the first release of IMT-Advanced, those two being LTE from release 10 ("LTE-Advanced") and WirelessMAN-Advanced [6] based on the IEEE 802.16m specification. The two can be viewed as the "family" of IMT-Advanced technologies as shown in Figure 1.4. The main IMT-Advanced recommendation, identifying the IMT-Advanced radio interface specifications, is

Copyright Elsevier 2024

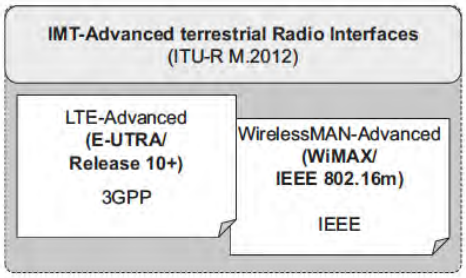


FIGURE 1.4

Radio interface technologies IMT-Advanced

ITU-R M.2012 [7]. As for the corresponding IMT-2000 specification, it contains an overview of each radio interface, followed by a list of references to the detailed specifications.

Looking further into the future, a new recommendation, "Framework and overall objectives of future development of IMT for 2020 and beyond" [120], is initiated within ITU-R WP 5D. The recommendation is a first step for defining developments of IMT in the future, looking at the future roles of IMT and how it can serve society, looking at market, user and technology trends, and spectrum implications. As a parallel activity, ITU-R WP5D is also developing a report on "Future technology trends of terrestrial IMT systems" [121], with a focus on the time period 2015-2020. The report will cover trends of future IMT technology aspects by looking at the technical and operational characteristics of IMT systems and how they are improved with the evolution of IMT technologies.

In this way, the report on technology trends relates the coming releases of LTE-Advanced in release 12 and beyond, while the recommendation on a future IMT Vision looks further ahead and beyond 2020. The deliverables will be ready by 2015 and will be available as input to the World Radio Conference 2015 (WRC'15, see next section). Chapter 21 discusses some of the technology components considered for LTE in release 12 and beyond, as well as future radio access in general.

1.3.2 Spectrum for IMT systems

Another major activity within ITU-R concerning IMT-Advanced has been to identify globally available spectrum, suitable for IMT systems. The spectrum work has involved sharing studies between IMT and other technologies in those bands. Adequate spectrum availability and globally harmonized spectrum are identified as essential for IMT-Advanced.

Spectrum for 3G was first identified at the *World Administrative Radio Congress* WARC-92, where 230 MHz was identified as intended for use by national administrations that want to implement IMT-2000. The so-called IMT-2000 "core band" at 2 GHz is in this frequency range and was the first band where 3G systems were deployed.

Copyright Elsevier 2024

Additional spectrum was identified for IMT-2000 at later World Radio communication conferences. WRC-2000 identified the existing 2G bands at 800/900 MHz and 1800/1900 MHz, plus an additional 190 MHz of spectrum at 2.6 GHz, all for IMT-2000. As additional spectrum for IMT-2000, WRC'07 identified a band at 450 MHz, the so-called "digital dividend" at 698–806 MHz, plus an additional 300 MHz of spectrum at higher frequencies. The applicability of these new bands varies on a regional and national basis. WRC'12 did not identify any additional spectrum allocations for IMT, but the issue was put on the agenda for WRC'15. It was also determined to study the use of the band 694-790 MHz for mobile services in Region 1 (Europe, Middle East, and Africa).

The worldwide frequency arrangements for IMT are outlined in ITU-R recommendation M.1036 [8], which is presently being updated with the arrangements for the most recent frequency bands added. The recommendation outlines the regional variations in how the bands are implemented and also identifies which parts of the spectrum are paired and which are unpaired. For the paired spectrum, the bands for uplink (mobile transmit) and downlink (base-station transmit) are identified for *Frequency-Division Duplex* (FDD) operation. The unpaired bands can, for example, be used for *Time-Division Duplex* (TDD) operation. Note that the band that is most globally deployed for 3G is still 2 GHz.

The same bands that were originally defined for IMT-2000 and used for 3G deployment are also being used for 4G, including LTE-Advanced deployment. The latest version of the ITU-R spectrum recommendation M.1036 [8] is renamed with a more generic title having the identifier "IMT" instead of "IMT-2000" and is now applicable for both LTE and LTE-Advanced deployments. Some regions and countries have issued licenses for new spectrum identified at WRC-2000 and WRC'07 with the intention of allowing 4G deployments, but in most cases new spectrum is licensed on a technology-neutral basis. Many new deployments are either 4G or 3G, but some are also a mix of 4G and 3G, sometimes also together with significant parts of the spectrum still used for 2G.

1.4 Drivers for LTE and LTE-Advanced

The evolution of 3G systems into 4G is driven by the creation and development of new services for mobile devices, and is enabled by advancement of the technology available for mobile systems. There has also been an evolution of the environment in which mobile systems are deployed and operated, in terms of competition between mobile operators, challenges from other mobile technologies, and new regulation of spectrum use and market aspects of mobile systems.

The rapid evolution of the technology used in telecommunication systems, consumer electronics, and specifically mobile devices has been remarkable in the last 20 years. Moore's law illustrates this and indicates a continuing evolution of processor performance and increased memory size, often combined with reduced size, power consumption, and cost for devices. Combined with a high-speed internet backbone often based on optical fiber

Copyright Elsevier 2024

networks, we see that a range of technology enablers are in place to go hand-in-hand with advancement in mobile communications technology such as LTE.

The rapid increase in use of the internet to provide all kinds of services since the 1990s started at the same time as 2G and 3G mobile systems came into widespread use. The natural next step was that those internet-based services also moved to the mobile devices, creating what is today know as *mobile broadband*. Being able to support the same *Internet Protocol* (IP)-based services in a mobile device that people use at home with a fixed broadband connection is a major challenge and a prime driver for the evolution of LTE. A few services were already supported by the evolved 2.5G systems, but it was not until the 3G and 4G systems were deployed, which are designed primarily for IP-based services, that the real mobile IP revolution took off. Increasing shares of mobile devices are "smartphones" that combine multimedia capabilities with applications that require mobile broadband data speeds, and a majority of mobile subscriptions will soon be for mobile broadband [112]. An interesting aspect of the migration of broadband services to mobile devices is that a mobile "flavor" is also added. The mobile position and the mobility and roaming capabilities do in fact create a whole new range of services tailored to the mobile environment.

Fixed telephony (POTS) and earlier generations of mobile technology were built for circuit switched services, primarily voice. The first data services over GSM were circuit switched, with packet-based GPRS coming in as a later addition. This also influenced the first development of 3G, which was based on circuit switched data, with packet-switched services as an add-on. It was not until the 3G evolution into HSPA and later LTE/LTE-Advanced that packet-switched services and IP were made the primary design target. The old circuit-switched services remain, but will on LTE be provided over IP, with Voice-over IP (VoIP) as an example.

IP is in itself service agnostic and thereby enables a range of services with different requirements. The main service-related design parameters for a radio interface supporting a variety of services are:

- **Data rate.** Many services with lower data rates such as voice services are important and still occupy a large part of a mobile network's overall capacity, but it is the higher data rate services that drive the design of the radio interface. The ever-increasing demand for higher data rates for mobile applications such as web browsing, streaming, and file transfer pushes the peak data rates for mobile systems from kbit/s for 2G to Mbit/s for 3G and getting close to Gbit/s for 4G.
- **Delay.** Interactive services such as real-time gaming, web browsing, and interactive file transfer have requirements for very low delay, making it a primary design target. There are, however, many applications such as email and television where the delay requirements are not as strict. The delay for a packet sent from a server to a client and back is called *latency.*
- **Capacity.** From the mobile system operator's point of view, it is not only the peak data rates provided to the end-user that are of importance, but also the total data rate that can be

Copyright Elsevier 2024

provided on average from each deployed base station site and per hertz of licensed spectrum. This measure of capacity is called *spectral efficiency*. In the case of capacity shortage in a mobile system, the Quality-of-Service (QoS) for the individual end-users may be degraded.

How these three main design parameters influenced the development of LTE is described in more detail in Chapter 7, and an evaluation of what performance is achieved for the design parameters previously discussed is presented in Chapter 20.

The demand for new services and for higher peak bit rates and system capacity is not only met by evolution of the technology to 4G. There is also a demand for more spectrum resources to expand systems and this demand also leads to more competition between an increasing number of mobile operators. In the initial deployment of 3G and during the development of 4G, there was also a competition between alternative technologies to provide mobile broadband services. While there are still systems such as CDMA2000 and WiMax in operation, the majority of 3G and 4G systems and the bulk of the mobile broadband growth are deployed with evolved 3G HSPA and 4G LTE-Advanced systems standardized in 3GPP.

With more spectrum coming into use for mobile broadband, there is a need to operate mobile systems in a number of different frequency bands, in spectrum allocations of different sizes, and sometimes also in fragmented spectrum. This calls for high spectrum flexibility with the possibility for a varying channel bandwidth, which was also a driver and an essential design parameter for LTE.

The demand for new mobile services and the evolution of the radio interface to LTE have served as drivers to evolve the core network. The core network developed for GSM in the 1980s was extended to support GPRS, EDGE, and WCDMA in the 1990s, but was still very much built around the circuit-switched domain. A *System Architecture Evolution* (SAE) was initiated at the same time as LTE development started and has resulted in an *Evolved Packet Core* (EPC), developed to support HSPA and LTE/LTE-Advanced, focusing on the packet-switched domain. For more details on SAE/EPC, please refer to [9].

1.5 Standardization of LTE

With a framework for IMT systems set up by the ITU-R, with spectrum made available by the WRC and with an ever-increasing demand for better performance, the task of specifying the LTE system that meets the design targets falls on 3GPP. 3GPP writes specifications for 2G, 3G, and 4G mobile systems, and 3GPP technologies are the most widely deployed in the world, with close to 90% of the world's 6.4 billion mobile subscriptions in Q1 2013 [112]. In order to understand how 3GPP works, it is important to understand the process of writing standards.

1.5.1 The standardization process

Setting a standard for mobile communication is not a one-time job; it is an ongoing process. The standardization fora are constantly evolving their standards trying to meet new demands

Copyright Elsevier 2024

for services and features. The standardization process is different in the different fora, but typically includes the four phases illustrated in Figure 1.5:

1. *Requirements*, where it is decided what is to be achieved by the standard.
2. *Architecture*, where the main building blocks and interfaces are decided.
3. *Detailed specifications*, where every interface is specified in detail.
4. *Testing and verification*, where the interface specifications are proven to work with real-life equipment.

These phases are overlapping and iterative. As an example, requirements can be added, changed, or dropped during the later phases if the technical solutions call for it. Likewise, the technical solution in the detailed specifications can change due to problems found in the testing and verification phase.

Standardization starts with the *requirements* phase, where the standards body decides what should be achieved with the standard. This phase is usually relatively short.

In the *architecture* phase, the standards body decides about the architecture—that is, the principles of how to meet the requirements. The architecture phase includes decisions about reference points and interfaces to be standardized. This phase is usually quite long and may change the requirements.

After the architecture phase, the *detailed specifications* phase starts. It is in this phase that the details for each of the identified interfaces are specified. During the detailed specification of the interfaces, the standards body may find that previous decisions in the architecture or even in the requirements phases need to be revisited.

Finally, the *testing and verification* phase starts. It is usually not a part of the actual standardization in the standards bodies, but takes place in parallel through testing by vendors and interoperability testing between vendors. This phase is the final proof of the standard. During the testing and verification phase, errors in the standard may still be found and those errors may change decisions in the detailed standard. Albeit not common, changes may also need to be made to the architecture or the requirements. To verify the standard, products are needed. Hence, the implementation of the products starts after (or during) the detailed specifications phase. The testing and verification phase ends when there are stable test specifications that can be used to verify that the equipment is fulfilling the standard.

Normally, it takes one to two years from the time when the standard is completed until commercial products are out on the market. However, if the standard is built from scratch, it may take longer since there are no stable components to build from.

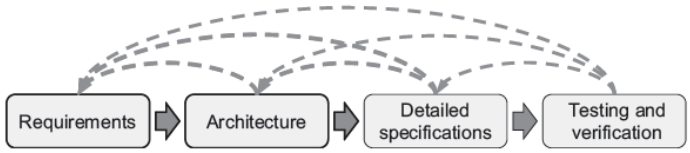


FIGURE 1.5

The standardization phases and iterative process

Copyright Elsevier 2024

1.5.2 The 3GPP process

The *Third-Generation Partnership Project* (3GPP) is the standards-developing body that specifies the LTE/LTE-Advanced, as well as the 3G UTRA and 2G GSM systems. 3GPP is a partnership project formed by the standards bodies ETSI, ARIB, TTC, TTA, CCSA, and ATIS. 3GPP consists of four *Technical Specifications Groups* (TSGs)—see Figure 1.6.

3GPP TSG RAN (*Radio Access Network*) is the technical specification group that has developed WCDMA, its evolution HSPA, as well as LTE/LTE-Advanced, and is in the forefront of mobile broadband technology. TSG RAN consists of five working groups (WGs):

1. RAN WG1, dealing with the physical layer specifications.
2. RAN WG2, dealing with the layer 2 and layer 3 radio interface specifications.
3. RAN WG3, dealing with the fixed RAN interfaces—for example, interfaces between nodes in the RAN—but also the interface between the RAN and the core network.
4. RAN WG4, dealing with the *radio frequency* (RF) and *radio resource management* (RRM) performance requirements.
5. RAN WG5, dealing with the terminal conformance testing.

The scope of 3GPP when it was formed in 1998 was to produce global specifications for a 3G mobile system based on an evolved GSM core network, including the WCDMA-based

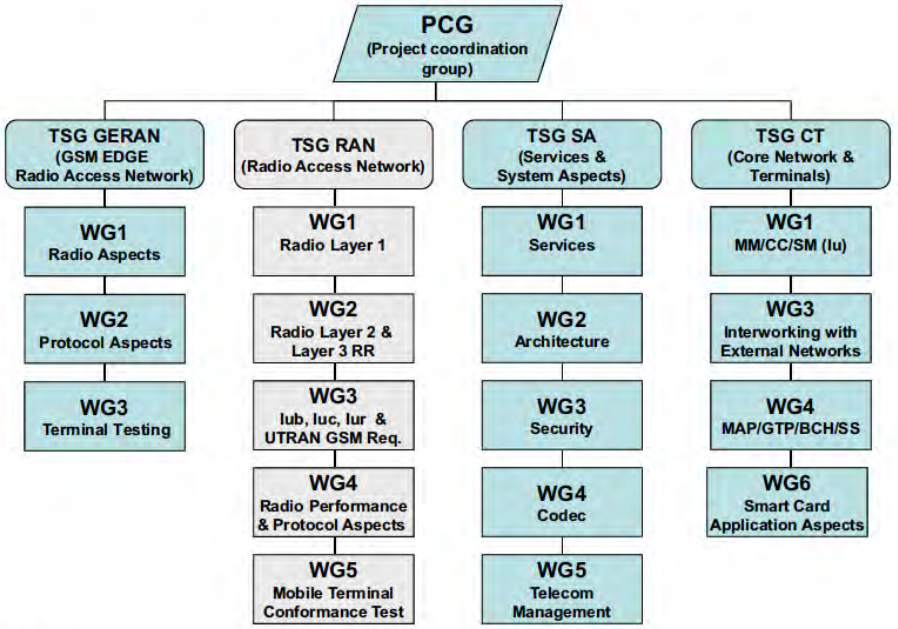


FIGURE 1.6

3GPP organization

Copyright Elsevier 2024

radio access of the UTRA FDD and the TD-CDMA-based radio access of the UTRA TDD mode. The task of developing and maintaining the GSM/EDGE specifications was added to 3GPP at a later stage and the work now also includes LTE (E-UTRA). The UTRA, E-UTRA and GSM/EDGE specifications are developed, maintained, and approved in 3GPP. After approval, the organizational partners transpose them into appropriate deliverables as standards in each region.

In parallel with the initial 3GPP work, a 3G system based on TD-SCDMA was developed in China. TD-SCDMA was eventually merged into release 4 of the 3GPP specifications as an additional TDD mode.

The work in 3GPP is carried out with relevant ITU recommendations in mind and the result of the work is also submitted to ITU as being part of IMT-2000 and IMT-Advanced. The organizational partners are obliged to identify regional requirements that may lead to options in the standard. Examples are regional frequency bands and special protection requirements local to a region. The specifications are developed with global roaming and circulation of terminals in mind. This implies that many regional requirements in essence will be global requirements for all terminals, since a roaming terminal has to meet the strictest of all regional requirements. Regional options in the specifications are thus more common for base stations than for terminals.

The specifications of all releases can be updated after each set of TSG meetings, which occur four times a year. The 3GPP documents are divided into releases, where each release has a set of features added compared to the previous release. The features are defined in Work Items agreed and undertaken by the TSGs. The releases from release 8 and onwards, with some main features listed for LTE, are shown in Figure 1.7. The date shown for each release is the day the content of the release was frozen. Release 10 of LTE is the first version

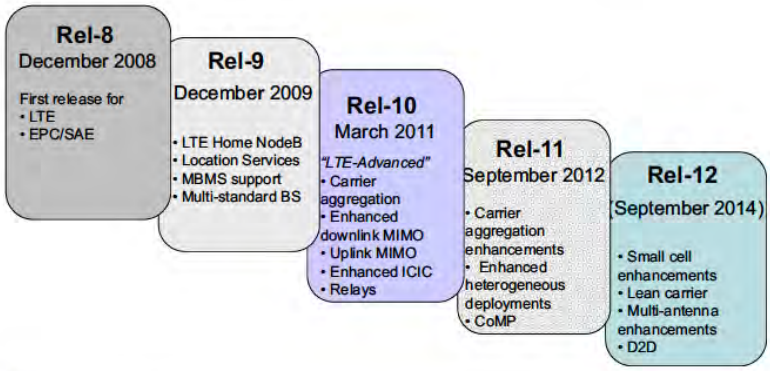


FIGURE 1.7

Releases of 3GPP specifications for LTE

Copyright Elsevier 2024

approved by ITU-R as an IMT-Advanced technology and is therefore also the first release named *LTE-Advanced*.

1.5.3 The 3G evolution to 4G

The first release of WCDMA Radio Access developed in TSG RAN was called release 99[3] and contained all features needed to meet the IMT-2000 requirements as defined by the ITU. This included circuit-switched voice and video services, and data services over both packet-switched and circuit-switched bearers. The first major addition of radio access features to WCDMA was HSPA, which was added in release 5 with *High Speed Downlink Packet Access* (HSDPA) and release 6 with *Enhanced Uplink*. These two are together referred to as HSPA. With HSPA, UTRA went beyond the definition of a 3G mobile system and also encompassed broadband mobile data.

The 3G evolution continued in 2004, when a workshop was organized to initiate work on the 3GPP *Long-Term Evolution* (LTE) radio interface. During an initial study phase that continued until mid-2005, the requirements and design targets for LTE were developed and documented [10]. Most notable are the requirements on high data rate at the cell edge and the importance of low delay, in addition to the normal capacity and peak data rate requirements. Furthermore, spectrum flexibility and maximum commonality between FDD and TDD solutions are pronounced. After continuing extensive studies of different basic physical layer technologies, it was decided that the LTE radio access should be based on OFDM in the downlink and DFT-precoded OFDM in the uplink. TSG RAN and its working groups then worked on the LTE specifications and the specifications were approved in December 2007. Work has since then continued on LTE, with new features added in each release, as shown in Figure 1.7. Chapters 7–19 will go through the LTE radio interface in more detail.

The work on IMT-Advanced within ITU-R WP5D came in 2008 into a phase where the detailed requirements and process were announced through a circular letter [5]. Among other things, this triggered new activities in 3GPP, where a study of *LTE-Advanced* was started. The task was to define requirements and investigate and propose technology components to be part of LTE-Advanced. The work was turned into a Work Item in 2009 in order to develop the detailed specifications.

Within 3GPP, LTE-Advanced is seen as the next major step in the evolution of LTE. LTE-Advanced is therefore not a new technology; it is an evolutionary step in the continuing development of LTE. As shown in Figure 1.7, the features that form LTE-Advanced are part of the 3GPP LTE specifications from release 10. Wider bandwidth through aggregation of multiple carriers and evolved use of advanced antenna techniques in both uplink and downlink are the major components added in LTE release 10 to reach the IMT-Advanced targets.

[3]For historical reasons, the first 3GPP release is named after the year it was frozen (1999), while the following releases are numbered 4, 5, 6, etc.

Copyright Elsevier 2024

The work on LTE-Advanced within 3GPP was planned with the ITU-R time frame in mind. LTE-Advanced was submitted as a candidate to the ITU-R in 2009 and is now included in the set of radio interface technologies announced by ITU-R [6] in October 2010 as a part of the IMT-Advanced radio interface specifications. This is very much aligned with what was from the start stated as a goal for LTE, namely that *LTE should provide the starting point for a smooth transition to 4G (= IMT-Advanced) radio access*.

The evolution of LTE and LTE-Advanced continues with features added in the 3GPP specification releases following release 10. A high-level summary of the features in releases 11 and 12 are shown in Figure 1.7. Since LTE-Advanced is an integral part of 3GPP LTE release 10 and onwards, it is described in detail together with the corresponding components of LTE in Chapters 7–19 of this book. 3GPP release 12 is presently (September 2013) not completed and frozen in 3GPP, but an overview is given in Chapter 21, where an outlook is also made toward new features that are expected in future releases.

Copyright Elsevier 2024

This page intentionally left blank

Copyright Elsevier 202

CHAPTER

High Data Rates in Mobile Communication

2

As discussed in Chapter 1, one main target for the evolution of mobile communication is to provide the possibility for significantly higher end-user data rates compared to what is achievable with, for example, the first releases of the 3G standards. This includes the possibility for higher peak data rates but, as pointed out in the previous chapter, even more so the possibility for significantly higher data rates over the entire cell area, also including, for example, users at the cell edge. The initial part of this chapter will briefly discuss some of the more fundamental constraints that exist in terms of what data rates can actually be achieved in different scenarios. This will provide a background to discussions in the later part of the chapter, as well as in subsequent chapters, concerning different means to increase the achievable data rates in various mobile-communication scenarios.

2.1 High data rates: fundamental constraints

In [11], Shannon provided the basic theoretical tools needed to determine the maximum rate, also known as the *channel capacity*, by which information can be transferred over a given communication channel. Although relatively complicated in the general case, for the special case of communication over a channel, for example a radio link, only impaired by additive white Gaussian noise, the channel capacity C is given by the simple expression [12]:

$$C = BW \cdot \log_2\left(1 + \frac{S}{N}\right) \tag{2.1}$$

where BW is the bandwidth available for the communication, S denotes the received signal power, and N denotes the power of the white noise impairing the received signal.

Already from (2.1) it should be clear that the two fundamental factors limiting the achievable data rate are the available received signal power, or more generally the available signal-power-to-noise-power ratio S/N, and the available bandwidth. To further clarify how and when these factors limit the achievable data rate, assume communication with a certain information rate R. The received signal power can then be expressed as $S = E_b \cdot R$, where E_b is the received energy per information bit. Furthermore, the noise power can be expressed as $N = N_0 \cdot BW$, where N_0 is the constant noise power spectral density measured in W/Hz.

4G: LTE/LTE-Advanced for Mobile Broadband. http://dx.doi.org/10.1016/B978-0-12-419985-9.00002-7
Copyright © 2014 Erik Dahlman, Stefan Parkvall and Johan Sköld. Published by Elsevier Ltd. All rights reserved.

Copyright Elsevier 2024
Samsung Exhibit 1017, Page 43 of 538
Samsung Electronics Co., Ltd. v. Asus Technology Licensing Inc.
IPR2024-01217

Clearly, the information rate can never exceed the channel capacity. Together with the above expressions for the received signal power and noise power, this leads to the inequality:

$$R \leq C = BW \cdot \log_2\left(1 + \frac{S}{N}\right) = BW \cdot \log_2\left(1 + \frac{E_b \cdot R}{N_0 \cdot BW}\right) \tag{2.2}$$

or, by defining the radio-link *bandwidth utilization* γ = R/BW:

$$\gamma \leq \log_2\left(1 = \gamma \cdot \frac{E_b}{N_0}\right) \tag{2.3}$$

This inequality can be reformulated to provide a lower bound on the required received energy per information bit, normalized to the noise power density, for a given bandwidth utilization γ:

$$\frac{E_b}{N_0} \geq \min\left\{\frac{E_b}{N_0}\right\} = \frac{2^\gamma - 1}{\gamma} \tag{2.4}$$

The rightmost expression—the minimum required E_b/N_0 at the receiver as a function of the bandwidth utilization—is illustrated in Figure 2.1. As can be seen, for bandwidth utilizations significantly less than 1—that is, for information rates substantially smaller than the utilized bandwidth—the minimum required E_b/N_0 is relatively constant, regardless of γ. For a given noise power density, any increase of the information data rate then implies a similar relative increase in the minimum required signal power $S = E_b \cdot R$ at the receiver. On the other hand, for bandwidth utilizations larger than 1, the minimum required E_b/N_0 increases rapidly with γ. Thus, in the case of data rates of the same order as or larger than the communication

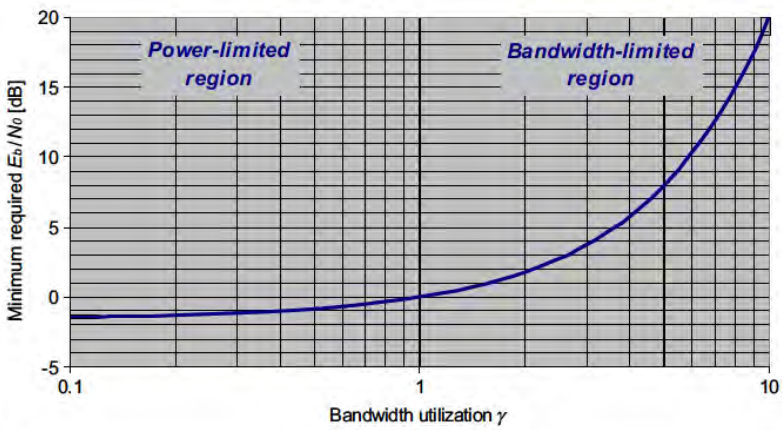


FIGURE 2.1

Minimum required E_b/N_0 at the receiver as a function of bandwidth utilization

Copyright Elsevier 2024

bandwidth, any further increase of the information data rate, without a corresponding increase in the available bandwidth, implies a larger, eventually much larger, relative increase in the minimum required received signal power.

2.1.1 High data rates in noise-limited scenarios

From the previous discussion, some basic conclusions can be drawn regarding the provisioning of higher data rates in a mobile-communication system when noise is the main source of radio-link impairment (a *noise-limited* scenario):

- The data rates that can be provided in such scenarios are always limited by the available received signal power or, in the general case, the received signal-power-to-noise-power ratio. Furthermore, any increase of the achievable data rate within a given bandwidth will require *at least* the same relative increase of the received signal power. At the same time, if sufficient received signal power can be made available, basically any data rate can, at least in theory, be provided within a given limited bandwidth.
- In the case of low-bandwidth utilization—that is, as long as the radio-link data rate is substantially lower than the available bandwidth—any further increase of the data rate requires *approximately the same* relative increase in the received signal power. This can be referred to as *power-limited* operation (in contrast to *bandwidth-limited* operation; see below) as, in this case, an increase in the available bandwidth does not substantially impact what received signal power is required for a certain data rate.
- On the other hand, in the case of high-bandwidth utilization—that is, in the case of data rates of the same order as or exceeding the available bandwidth—any further increase in the data rate requires a *much larger* relative increase in the received signal power unless the bandwidth is increased in proportion to the increase in data rate. This can be referred to as *bandwidth-limited operation* as, in this case, an increase in the bandwidth will reduce the received signal power required for a certain data rate.

Thus, to make efficient use of the available received signal power or, in the general case, the available signal-to-noise ratio, the transmission bandwidth should at least be of the same order as the data rates to be provided.

Assuming a constant transmit power, the received signal power can always be increased by reducing the distance between the transmitter and the receiver, thereby reducing the attenuation of the signal as it propagates from the transmitter to the receiver. In a noise-limited scenario it is, at least in theory, always possible to increase the achievable data rates, assuming that one is prepared to accept a reduction in the transmitter/receiver distance—that is, a reduced range. In a mobile-communication system this would correspond to a reduced cell size and the need for more cell sites to cover the same overall area. In particular, providing data rates of the same order as or larger than the available bandwidth—that is, with a high-bandwidth utilization—would require a significant cell-size reduction. Alternatively, one has to accept that the high data rates are only available for terminals in the center of the cell and not over the entire cell area.

Copyright Elsevier 2024

Another means to increase the overall received signal power for a given transmit power is the use of additional antennas at the receiver side, also known as *receive-antenna diversity*. Multiple receive antennas can be applied at the base station (that is, for the uplink) or at the terminal (that is, for the downlink). By proper combination of the signals received at the different antennas, the signal-to-noise ratio after the antenna combination can be increased in proportion to the number of receive antennas, thereby allowing for higher data rates for a given transmitter/receiver distance.

Multiple antennas can also be applied at the transmitter side, typically at the base station, and can be used to focus a given total transmit power in the direction of the receiver—that is, toward the target terminal. This will increase the received signal power and, once again, allow for higher data rates for a given transmitter/receiver distance.

However, providing higher data rates by the use of multiple transmit or receive antennas is only efficient up to a certain level—that is, as long as the data rates are power limited rather than bandwidth limited. Beyond this point, the achievable data rates start to saturate and any further increase in the number of transmit or receive antennas, although leading to a correspondingly improved signal-to-noise ratio at the receiver, will only provide a marginal increase in the achievable data rates. This saturation in achievable data rates can be avoided by the use of multiple antennas at both the transmitter *and* the receiver, enabling what can be referred to as *spatial multiplexing*, often also referred to as *MIMO* (Multiple-Input Multiple-Output). Different types of multi-antenna techniques, including spatial multiplexing, will be discussed in more detail in Chapter 5. Multi-antenna techniques for the specific case of LTE are discussed in Chapters 10 and 11.

An alternative to increasing the received signal power is to reduce the noise power, or more exactly the noise power density, at the receiver. This can, at least to some extent, be achieved by more advanced receiver RF design, allowing for a reduced receiver noise figure.

2.1.2 Higher data rates in interference-limited scenarios

The previous discussion assumed communication over a radio link only impaired by noise. However, in actual mobile-communication scenarios, interference from transmissions in neighboring cells, also referred to as *inter-cell interference*, is often the dominant source of radio-link impairment, more so than noise. This is especially the case in small-cell deployments with a high traffic load. Furthermore, in addition to inter-cell interference there may also be interference from other transmissions *within the current cell*, also referred to as *intra-cell interference*.

In many respects the impact of interference on a radio link is similar to that of noise. In particular, the basic principles discussed above apply also to a scenario where interference is the main radio-link impairment:

- The maximum data rate that can be achieved in a given bandwidth is limited by the available signal-power-to-interference-power ratio.
- Providing data rates larger than the available bandwidth (high-bandwidth utilization) is costly in the sense that it requires a disproportionately high signal-to-interference ratio.

Copyright Elsevier 2024

Also, similar to a scenario where noise is the dominant radio-link impairment, reducing the cell size as well as the use of multi-antenna techniques are key means to increase the achievable data rates in an interference-limited scenario:

- Reducing the cell size will obviously reduce the number of users, and thus also the overall traffic, per cell. This will reduce the relative interference level and allow for higher data rates.
- Similar to the increase in signal-to-noise ratio, proper combination of the signals received at multiple antennas will also increase the signal-to-interference ratio after the antenna combination.
- The use of beam-forming by means of multiple transmit antennas will focus the transmit power in the direction of the target receiver, leading to reduced interference to other radio links and improving the overall signal-to-interference ratio in the system.

One important difference between interference and noise is that interference, in contrast to noise, typically has a certain structure which makes it, at least to some extent, predictable and possible to further suppress or even remove completely. As an example, a dominant interfering signal may arrive from a certain direction, in which case the corresponding interference can be further suppressed, or even completely removed, by means of *spatial processing* using multiple antennas at the receiver. This will be further discussed in Chapter 5. Also, any differences in the spectral properties between the target signal and an interfering signal can be used to suppress the interferer and thus reduce the overall interference level.

2.2 Higher data rates within a limited bandwidth: higher-order modulation

As discussed in the previous section, providing data rates larger than the available bandwidth is fundamentally inefficient in the sense that it requires disproportionately high signal-to-noise and signal-to-interference ratios at the receiver. Still, bandwidth is often a scarce and expensive resource and, at least in some mobile-communication scenarios, high signal-to-noise and signal-to-interference ratios can be made available, for example in small-cell environments with a low traffic load or for terminals close to the cell site. Mobile-communication systems should preferably be designed to be able to take advantage of such scenarios—that is, they should be able to offer very high data rates within a limited bandwidth when the radio conditions so allow.

A straightforward means to provide higher data rates within a given transmission bandwidth is the use of *higher-order modulation*, implying that the modulation alphabet is extended to include additional signaling alternatives and thus allowing for more bits of information to be communicated per modulation symbol.

Copyright Elsevier 2024

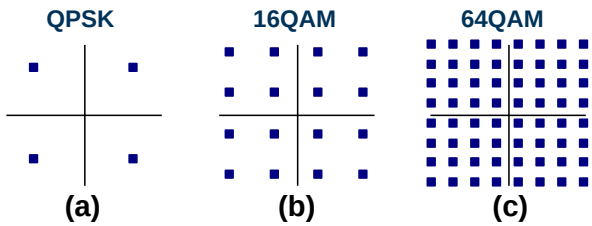


FIGURE 2.2

Signal constellations for (a) QPSK, (b) 16QAM, and (c) 64QAM

In the case of QPSK modulation, which is the modulation scheme used for the downlink in the first releases of the 3G mobile-communication standards (WCDMA and CDMA2000), the modulation alphabet consists of four different signaling alternatives. These four signaling alternatives can be illustrated as four different points in a two-dimensional plane (see Figure 2.2a). With four different signaling alternatives, QPSK allows for up to 2 bits of information to be communicated during each modulation-symbol interval. By extending to 16QAM modulation (Figure 2.2b), 16 different signaling alternatives are available. The use of 16QAM thus allows for up to 4 bits of information to be communicated per symbol interval. Further extension to 64QAM (Figure 2.2c), with 64 different signaling alternatives, allows for up to 6 bits of information to be communicated per symbol interval. At the same time, the bandwidth of the transmitted signal is, at least in principle, independent of the size of the modulation alphabet and mainly depends on the modulation rate (the number of modulation symbols per second). The maximum bandwidth utilization, expressed in bits/s/Hz, of 16QAM and 64QAM are, at least in principle, two and three times that of QPSK respectively.

It should be pointed out that there are many other possible modulation schemes, in addition to those illustrated in Figure 2.2. One example is 8PSK, consisting of eight signaling alternatives and providing up to 3 bits of information per modulation symbol. Readers are referred to [12] for a more thorough discussion on different modulation schemes.

The use of higher-order modulation provides the possibility for higher bandwidth utilization—that is, the possibility to provide higher data rates within a given bandwidth. However, the higher bandwidth utilization comes at the cost of reduced robustness to noise and interference. Alternatively expressed, higher-order modulation schemes, such as 16QAM or 64QAM, require a higher E_b/N_0 at the receiver for a given bit-error probability, compared to QPSK. This is in line with the discussion in the previous section, where it was concluded that high-bandwidth utilization—a high information rate within a limited bandwidth—in general requires a higher receiver E_b/N_0.

2.2.1 Higher-order modulation in combination with channel coding

Higher-order modulation schemes such as 16QAM and 64QAM require a higher receiver E_b/N_0 for a given error rate, compared to QPSK. However, in combination with channel

Copyright Elsevier 2024

coding the use of higher-order modulation will sometimes be more efficient—that is, require a lower receiver E_b/N_0 for a given error rate—compared to the use of lower-order modulation such as QPSK. This may, for example, occur when the target bandwidth utilization implies that, with lower-order modulation, no or very little channel coding can be applied. In such a case, the additional channel coding that can be applied by using a higher-order modulation scheme such as 16QAM may lead to an overall gain in power efficiency compared to the use of QPSK.

As an example, if a bandwidth utilization of close to 2 information bits per modulation symbol is required, QPSK modulation would allow for very limited channel coding (channel-coding rate close to 1). On the other hand, the use of 16QAM modulation would allow for a channel-coding rate of the order of one-half. Similarly, if a bandwidth efficiency close to 4 information bits per modulation symbol is required, the use of 64QAM may be more efficient than 16QAM modulation, taking into account the possibility for lower-rate channel coding and corresponding additional coding gain in the case of 64QAM. It should be noted that this does not contradict the general discussion in Section 2.1 where it was concluded that transmission with high-bandwidth utilization is inherently power inefficient. The use of rate 1/2 channel coding for 16QAM reduces the information data rate, and thus also the bandwidth utilization, to the same level as uncoded QPSK.

From the discussion above it can be concluded that, for a given signal-to-noise/interference ratio, a certain combination of modulation scheme and channel-coding rate is optimal in the sense that it can deliver the highest-bandwidth utilization (the highest data rate within a given bandwidth) for that signal-to-noise/interference ratio.

2.2.2 Variations in instantaneous transmit power

A general drawback of higher-order modulation schemes such as 16QAM and 64QAM, where information is also encoded in the instantaneous amplitude of the modulated signal, is that the modulated signal will have larger variations, and larger peaks, in its instantaneous power. This can be seen in Figure 2.3, which illustrates the distribution of the instantaneous power, more specifically the probability that the instantaneous power is above a certain value, for QPSK, 16QAM, and 64QAM respectively. The probability for large peaks in the instantaneous power is higher in the case of higher-order modulation.

Larger peaks in the instantaneous signal power imply that the transmitter power amplifier must be over-dimensioned to avoid power-amplifier nonlinearities, occurring at high instantaneous power levels, causing corruption to the signal to be transmitted. As a consequence, the power-amplifier efficiency will be reduced, leading to increased power consumption. In addition, there will be a negative impact on the power-amplifier cost. Alternatively, the average transmit power must be reduced, implying a reduced range for a given data rate. High power-amplifier efficiency is especially important for the terminal—that is, in the uplink direction—due to the importance of low mobile-terminal power consumption and cost. For the base station, high power-amplifier efficiency, although far from irrelevant, is

Copyright Elsevier 2024

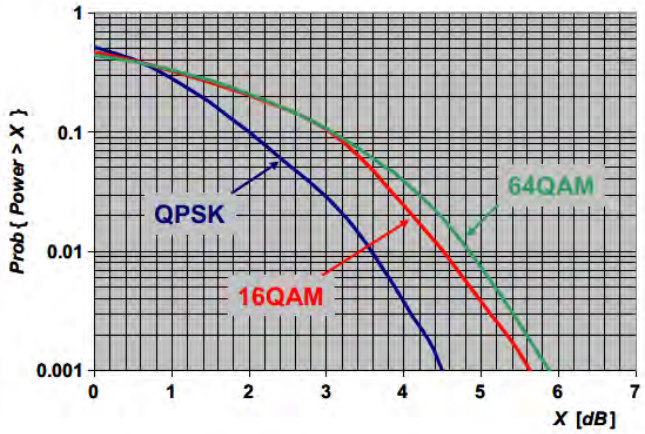


FIGURE 2.3

Distribution of instantaneous power for different modulation schemes. Average power is the same in all cases

still somewhat less important. Thus, large peaks in the instantaneous signal power are less of an issue for the downlink compared to the uplink and, consequently, higher-order modulation is more suitable for the downlink compared to the uplink.

2.3 Wider bandwidth including multi-carrier transmission

As was shown in Section 2.1, transmission with a high-bandwidth utilization is fundamentally power inefficient in the sense that it will require disproportionately high signal-to-noise and signal-to-interference ratios for a given data rate. Providing very high data rates within a limited bandwidth, for example by means of higher-order modulation, is thus only possible in situations where relatively high signal-to-noise and signal-to-interference ratios can be made available, for example in small-cell environments with low traffic load or for terminals close to the cell site.

Instead, to provide high data rates as efficiently as possible in terms of required signal-to-noise and signal-to-interference ratios, implying as good coverage as possible for high data rates, the transmission bandwidth should be at least of the same order as the data rates to be provided. Having in mind that the provisioning of higher data rates with good coverage is one of the main targets for mobile communication, it can thus be concluded that support for even wider transmission bandwidth is an important part of this evolution.

Copyright Elsevier 2024

However, there are several critical issues related to the use of wider transmission bandwidths in a mobile-communication system:

- Spectrum is, as already mentioned, often a scarce and expensive resource, and it may be difficult to find spectrum allocations of sufficient size to allow for very wideband transmission, especially at lower-frequency bands.
- The use of wider transmission and reception bandwidths has an impact on the complexity of the radio equipment, both at the base station and at the terminal. As an example, a wider transmission bandwidth has a direct impact on the transmitter and the receiver sampling rates, and thus on the complexity and power consumption of digital-to-analog and analog-to-digital converters, as well as front-end digital signal processing. RF components are also, in general, more complicated to design and more expensive to produce, the wider the bandwidth they have to handle.

These two issues are mainly outside the scope of this book. However, a more specific technical issue related to wider-band transmission is the increased corruption of the transmitted signal due to time dispersion on the radio channel. Time dispersion occurs when the transmitted signal propagates to the receiver via multiple paths with different delays (see Figure 2.4a). In the frequency domain, a time-dispersive channel corresponds to a non-constant channel frequency response, as illustrated in Figure 2.4b. This radio-channel *frequency selectivity* will corrupt the frequency-domain structure of the transmitted signal and lead to higher error rates for given signal-to-noise/interference ratios. Every radio channel is subject to frequency selectivity, at least to some extent. However, the extent to which the frequency selectivity impacts the radio communication depends on the bandwidth of the transmitted signal with, in general, larger impact for wider-band transmission. The amount of radio-channel frequency selectivity also depends on the environment, with typically less frequency selectivity (less time dispersion) in the case of small cells and in environments with few obstructions and potential reflectors, such as rural environments.

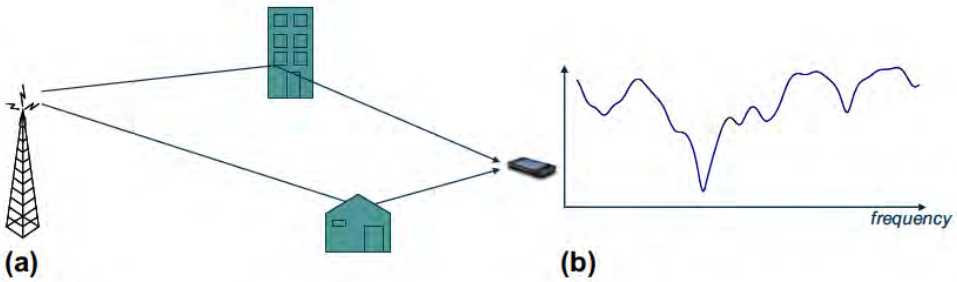


FIGURE 2.4

Multi-path propagation causing time dispersion and radio-channel frequency selectivity

Copyright Elsevier 2024

It should be noted that Figure 2.4b illustrates a "snapshot" of the channel frequency response. As a terminal is moving through the environment, the detailed structure of the multi-path propagation, and thus also the detailed structure of the channel frequency response, may vary rapidly with time. The rate of the variations in the channel frequency response is related to the channel *Doppler spread*, f_D, defined as $f_D = v/c \cdot f_c$, where v is the speed of the terminal, f_c is the carrier frequency (for example, 2 GHz), and c is the speed of light.

Receiver-side *equalization* [12] has for many years been used to counteract signal corruption due to radio-channel frequency selectivity. Equalization has been shown to provide satisfactory performance with reasonable complexity at least up to bandwidths corresponding to the WCDMA bandwidth of 5 MHz (see, for example, [13]). However, if the transmission bandwidth is further increased up to, for example, 20 MHz, as is done in LTE, the complexity of straightforward high-performance equalization starts to become a serious issue. One option is to apply less optimal equalization, with a corresponding negative impact on the equalizer capability to counteract the signal corruption due to radio-channel frequency selectivity and thus a corresponding negative impact on the radio-link performance.

An alternative approach is to consider specific transmission schemes and signal designs that allow for good radio-link performance also in the case of substantial radio-channel frequency selectivity without a prohibitively large receiver complexity. In the following, two such approaches to wider-band transmission will be discussed:

- The use of different types of *multi-carrier transmission*—that is, transmitting an overall wider-band signal as several more narrowband frequency-multiplexed signals (see below). One special case of multi-carrier transmission is *OFDM transmission*, to be discussed in more detail in Chapter 3.
- The use of specific *single-carrier* transmission schemes, especially designed to allow for efficient but still reasonably low-complexity equalization. This is further discussed in Chapter 4.

2.3.1 Multi-carrier transmission

One way to increase the overall transmission bandwidth, without suffering from increased signal corruption due to radio-channel frequency selectivity, is the use of so-called *multi-carrier transmission*. As illustrated in Figure 2.5, multi-carrier transmission implies that, instead of transmitting a single more wideband signal, multiple more narrowband signals, often referred to as *subcarriers*, are frequency multiplexed and jointly transmitted over the same radio link to the same receiver. By transmitting M signals in parallel over the same radio link, the overall data rate can be increased up to M times. At the same time, the impact in terms of signal corruption due to radio-channel frequency selectivity depends on the bandwidth of each subcarrier. Thus, the impact from a frequency-selective channel is essentially the same as for a more narrowband transmission scheme with a bandwidth that corresponds to the bandwidth of each subcarrier.

A drawback of the kind of multi-carrier evolution outlined in Figure 2.5, where an existing more narrowband radio-access technology is extended to a wider overall transmission bandwidth

Copyright Elsevier 2024

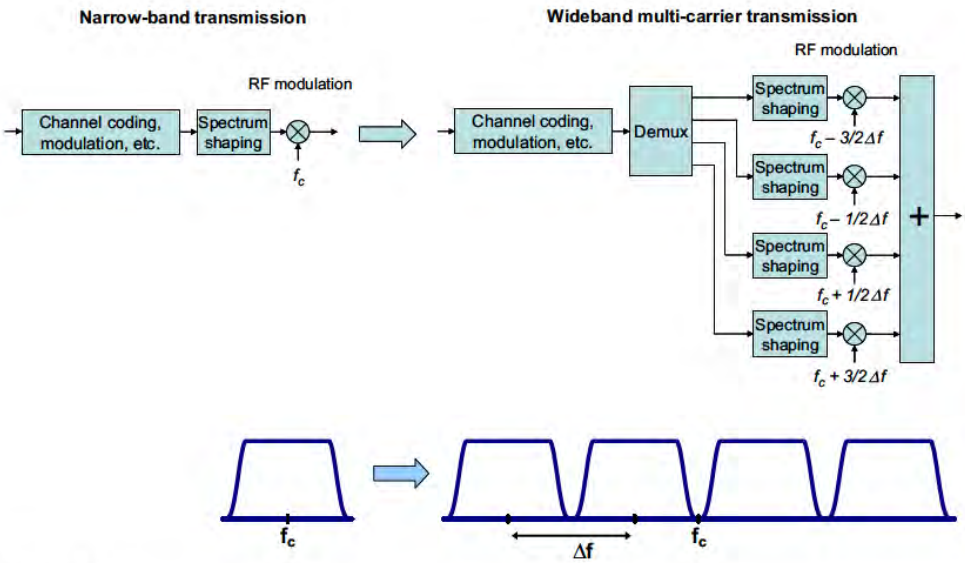


FIGURE 2.5

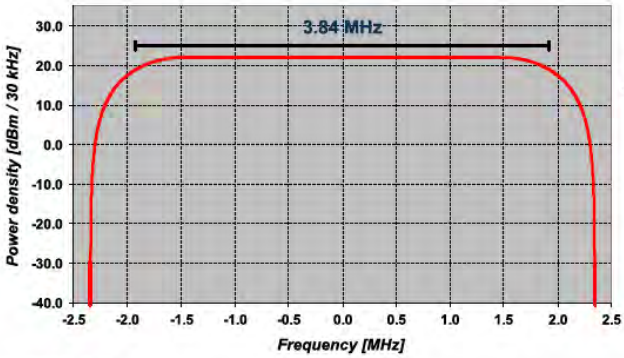
Extension to wider transmission bandwidth by means of multi-carrier transmission

by the parallel transmission of M more narrowband carriers, is that the spectrum of each subcarrier typically does not allow for very tight subcarrier "packing." This is illustrated by the "valleys" in the overall multi-carrier spectrum outlined in the lower part of Figure 2.5. This has a somewhat negative impact on the overall bandwidth efficiency of this kind of multi-carrier transmission.

As an example, consider the multi-carrier evolution of WCDMA/HSPA toward wider bandwidth. WCDMA has a modulation rate, also referred to as the *WCDMA chip rate*, of $f_{cr} = 3.84$ Mchips/s. However, due to spectrum shaping, even the theoretical WCDMA spectrum, not including spectrum widening due to transmitter imperfections, has a bandwidth that significantly exceeds 3.84 MHz. More specifically, as can be seen in Figure 2.6, the theoretical WCDMA spectrum has a *raised-cosine* shape with roll-off $\alpha = 0.22$. As a consequence, the bandwidth outside of which the WCDMA theoretical spectrum equals zero is approximately 4.7 MHz (see right part of Figure 2.6).

For a straightforward multi-carrier extension of WCDMA, the subcarriers must thus be spaced approximately 4.7 MHz from each other to completely avoid inter-subcarrier interference. It should be noted, though, that a smaller subcarrier spacing can be used with only limited inter-subcarrier interference.

Copyright Elsevier 2024



$$P(f)=\begin{cases}1 & |f|<f_1\\ \frac{1}{2}\left(1+\cos\left(\frac{\pi(|f|-f_1)}{f_{cr}\alpha}\right)\right) & f_1\le|f|<f_2\\ 0 & |f|>f_2\end{cases}$$

$$f_1=f_{cr}\cdot\frac{(1-\alpha)}{2}\qquad f_2=f_{cr}\cdot\frac{(1+\alpha)}{2}$$

$$BW=2\cdot f_2=f_{cr}\cdot(1-\alpha)\approx 4.7MHz$$

FIGURE 2.6

Theoretical WCDMA spectrum. Raised-cosine shape with roll-off $\alpha = 0.22$

A second drawback of multi-carrier transmission is that, similar to the use of higher-order modulation, the parallel transmission of multiple carriers will lead to larger variations in the instantaneous transmit power. Thus, similar to the use of higher-order modulation, multi-carrier transmission will have a negative impact on the transmitter power-amplifier efficiency, implying increased transmitter power consumption and increased power-amplifier cost. Alternatively, the average transmit power must be reduced, implying a reduced range for a given data rate. For this reason, similar to the use of higher-order modulation, multi-carrier transmission is more suitable for the downlink (base-station transmission), compared to the uplink (mobile-terminal transmission), due to the higher importance of high power-amplifier efficiency at the terminal.

The main advantage with the kind of multi-carrier extension outlined in Figure 2.5 is that it provides a very smooth evolution, in terms of both radio equipment and spectrum, of an already existing radio-access technology to wider transmission bandwidth and a corresponding possibility for higher data rates, especially for the downlink. In essence this kind of multi-carrier evolution to wider bandwidth can be designed so that, for legacy terminals not capable of multi-carrier reception, each downlink "subcarrier" will appear as an original, more narrowband carrier, while, for a multi-carrier-capable terminal, the network can make use of the full multi-carrier bandwidth to provide higher data rates.

The next chapter will discuss, in more detail, a different approach to multi-carrier transmission, based on the so-called OFDM technique.

Copyright Elsevier 2024

CHAPTER

OFDM Transmission

3

In this chapter, a more detailed overview of OFDM or *Orthogonal Frequency-Division Multiplexing* will be given. OFDM is the transmission scheme used for 3GPP LTE and is also used for several other radio-access technologies, for example WiMAX [14] and the DVB broadcast technologies [15].

3.1 Basic principles of OFDM

Transmission by means of OFDM can be seen as a kind of multi-carrier transmission. The basic characteristics of OFDM transmission, which distinguish it from a straightforward multi-carrier extension of a more narrowband transmission scheme as outlined in Figure 2.5 in the previous chapter, are:

- The use of a typically very large number of relatively narrowband subcarriers. In contrast, a straightforward multi-carrier extension as outlined in Figure 2.5 would typically consist of only a few subcarriers, each with a relatively wide bandwidth. As an example, the HSPA multi-carrier evolution to a 20 MHz overall transmission bandwidth consists of four (sub)carriers, each with a bandwidth of the order of 5 MHz. In comparison, OFDM transmission may imply that several hundred subcarriers are transmitted over the same radio link to the same receiver.
- Simple rectangular pulse shaping as illustrated in Figure 3.1a. This corresponds to a sinc-square-shaped per-subcarrier spectrum, as illustrated in Figure 3.1b.
- Tight frequency-domain packing of the subcarriers with a subcarrier spacing $\Delta f = 1/T_u$, where T_u is the per-subcarrier modulation-symbol time (see Figure 3.2). The subcarrier spacing is thus equal to the per-subcarrier modulation rate $1/T_u$.[1]

An illustrative description of a basic OFDM modulator is provided in Figure 3.3. It consists of a bank of N_c complex modulators, where each modulator corresponds to one OFDM subcarrier.

In complex baseband notation, a basic OFDM signal $x(t)$ during the time interval $mT_u \leq t < (m+1)T_u$ can thus be expressed as:

$$x(t) = \sum_{k=0}^{N_c - 1} x_k(t) = \sum_{k=0}^{N_c - 1} a_k^{(m)} e^{j2\pi k \Delta f t} \tag{3.1}$$

[1]This ignores a possible *cyclic prefix*, see Section 3.4.

4G: LTE/LTE-Advanced for Mobile Broadband. http://dx.doi.org/10.1016/B978-0-12-419985-9.00003-9
Copyright © 2014 Erik Dahlman, Stefan Parkvall and Johan Sköld. Published by Elsevier Ltd. All rights reserved.

Copyright Elsevier 2024
Samsung Exhibit 1017, Page 55 of 538
Samsung Electronics Co., Ltd. v. Asus Technology Licensing Inc.
IPR2024-01217

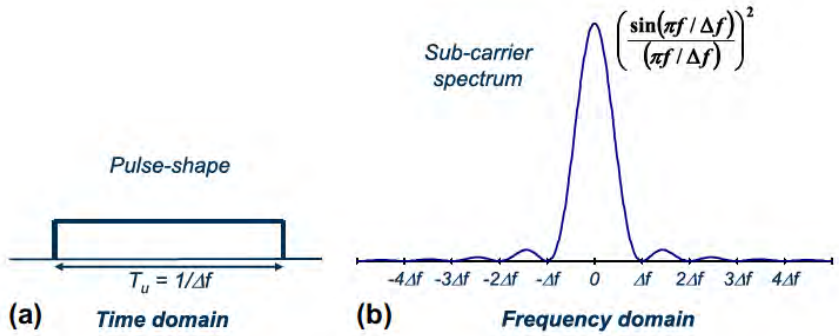


FIGURE 3.1

(a) Per-subcarrier pulse shape and (b) spectrum for basic OFDM transmission

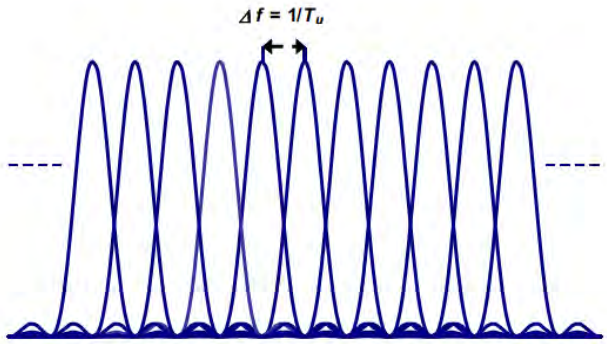


FIGURE 3.2

OFDM subcarrier spacing

where $x_k(t)$ is the kth modulated subcarrier with frequency $f_k = k \cdot \Delta f$ and $a_k^{(m)}$ is the, generally complex, modulation symbol applied to the kth subcarrier during the mth OFDM symbol interval—that is, during the time interval $mT_u \leq t < (m+1)T_u$. OFDM transmission is block based, implying that, during each OFDM symbol interval, N_c modulation symbols are transmitted in parallel. The modulation symbols can be from any modulation alphabet, such as QPSK, 16QAM, or 64QAM.

The number of OFDM subcarriers can range from less than one hundred to several thousand, with the subcarrier spacing ranging from several hundred kHz down to a few kHz. What subcarrier spacing to use depends on what types of environments the system is to operate in, including such aspects as the maximum expected radio-channel frequency selectivity (maximum expected time dispersion) and the maximum expected rate of channel variations (maximum expected Doppler spread). Once the subcarrier spacing has been selected, the number of subcarriers can be decided based on the assumed overall transmission bandwidth,

Copyright Elsevier 2024

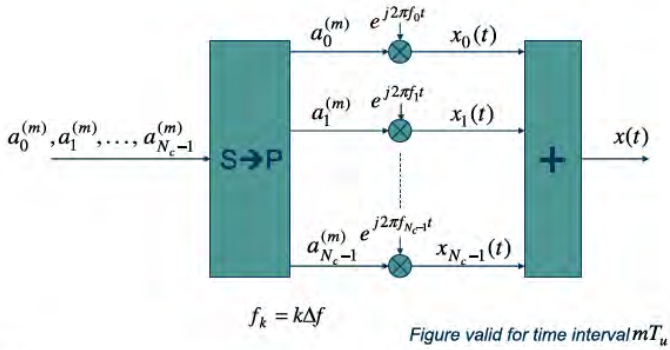


FIGURE 3.3

OFDM modulation

taking into account acceptable out-of-band emission, etc. The selection of OFDM subcarrier spacing and number of subcarriers is discussed in somewhat more detail in Section 3.8.

As an example, for 3GPP LTE the basic subcarrier spacing equals 15 kHz. On the other hand, the number of subcarriers depends on the transmission bandwidth, with 600 subcarriers in the case of operation in a 10 MHz spectrum allocation and correspondingly fewer/more subcarriers in the case of smaller/larger overall transmission bandwidths.

The term *Orthogonal Frequency-Division Multiplex* is due to the fact that two modulated OFDM subcarriers x_{k_1} and x_{k_2} are mutually *orthogonal* over the time interval $mT_u \le t < (m+1)T_u$, that is:

$$\int_{mT_u}^{(m+1)T_u} x_{k_1}(t) x_{k_2}^*(t)\, dt = \int_{mT_u}^{(m+1)T_u} a_{k_1} a_{k_2}^* e^{j2\pi k_1 \Delta f t} e^{j2\pi k_2 \Delta f t}\, dt = 0 \quad \text{for } k_1 \neq k_2 \tag{3.2}$$

Thus, basic OFDM transmission can be seen as the modulation of a set of orthogonal functions $\varphi_k(t)$, where:

$$\varphi_k(t) = \begin{cases} e^{j2\pi k \Delta f t} & 0 \le t < T_u \\ 0 & \text{otherwise} \end{cases} \tag{3.3}$$

The "physical resource" in the case of OFDM transmission is often illustrated as a time–frequency grid according to Figure 3.4, where each "column" corresponds to one OFDM symbol and each "row" corresponds to one OFDM subcarrier.

3.2 OFDM demodulation

Figure 3.5 illustrates the basic principle of OFDM demodulation consisting of a bank of correlators, one for each subcarrier. Taking into account the orthogonality between subcarriers according to (3.2), it is clear that, in the ideal case, two OFDM subcarriers do not cause any

Copyright Elsevier 2024

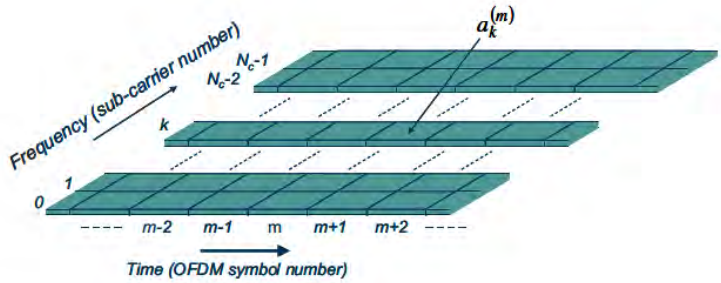


FIGURE 3.4

OFDM time frequency grid

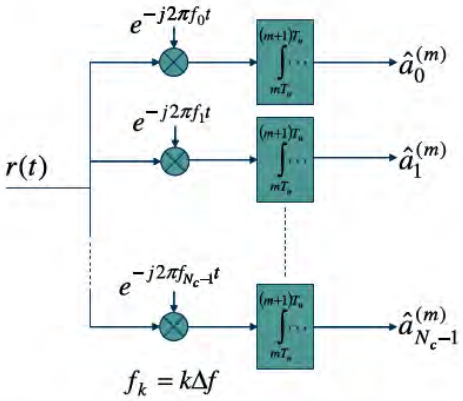


FIGURE 3.5

Basic principle of OFDM demodulation

interference to each other after demodulation. Note that this is the case despite the fact that the spectrum of neighbor subcarriers clearly overlaps, as can be seen from Figure 3.2. Thus, the avoidance of interference between OFDM subcarriers is not simply due to a subcarrier spectrum separation, which is, for example, the case for the kind of straightforward multi-carrier extension outlined in Figure 2.5 in the previous chapter. Rather, the subcarrier orthogonality is due to the *specific* frequency-domain structure of each subcarrier in combination with the *specific* choice of a subcarrier spacing Δf equal to the per-subcarrier symbol rate $1/T_u$. However, this also implies that, in contrast to the kind of multi-carrier transmission outlined in Section 2.3.1 of the previous chapter, any corruption of the frequency-domain structure of the OFDM subcarriers, for example due to a frequency-selective radio channel, may lead to a loss of inter-subcarrier orthogonality and thus to interference between subcarriers. To handle this and to

Copyright Elsevier 2024

make an OFDM signal truly robust to radio-channel frequency selectivity, *cyclic-prefix insertion* is typically used, as will be further discussed in Section 3.4.

3.3 OFDM implementation using IFFT/FFT processing

Although a bank of modulators/correlators according to Figures 3.3 and 3.5 can be used to illustrate the basic principles of OFDM modulation and demodulation respectively, these are not the most appropriate modulator/demodulator structures for actual implementation. Actually, due to its specific structure and the selection of a subcarrier spacing Δf equal to the per-subcarrier symbol rate $1/T_u$, OFDM allows for low-complexity implementation by means of computationally efficient *Fast Fourier Transform* (FFT) processing.

To confirm this, consider a time-discrete (sampled) OFDM signal where it is assumed that the sampling rate f_s is a multiple of the subcarrier spacing Δf that is —$f_s = 1/T_s = N \cdot \Delta f$. The parameter N should be chosen so that the sampling theorem [12] is sufficiently fulfilled.[2] As $N_c \cdot \Delta f$ can be seen as the nominal bandwidth of the OFDM signal, this implies that N should exceed N_c with a sufficient margin.

With these assumptions, the time-discrete OFDM signal can be expressed as:[3]

$$x_n = x(nT_s) = \sum_{k=0}^{N_c\ 1} a_k e^{j2\pi k\Delta f n T_s} = \sum_{k=0}^{N_c\ 1} a_k e^{j2\pi kn/N} = \sum_{k=0}^{N\ 1} a'_k e^{j2\pi kn/N} \tag{3.4}$$

where

$$a'_k = \begin{cases} a_k & 0 \le k < N_c \\ 0 & N_c \le k < N \end{cases} \tag{3.5}$$

Thus, the sequence x_n, or, in other words, the sampled OFDM signal, is the size-N Inverse Discrete Fourier Transform (IDFT) of the block of modulation symbols $a_0, a_1, ..., a_{N_c\ 1}$ extended with zeros to length N. OFDM modulation can be implemented by means of IDFT processing followed by digital-to-analog conversion, as illustrated in Figure 3.6. In particular, by selecting the IDFT size N equal to 2^m for some integer m, the OFDM modulation can be implemented by means of implementation-efficient radix-2 Inverse Fast Fourier Transform (IFFT) processing. It should be noted that the ratio N/N_c, which could be seen as the *over-sampling* of the time-discrete OFDM signal, could very well be, and typically is, a non-integer number. As an example and as already mentioned, for 3GPP LTE the number of subcarriers N_c is approximately 600 in the case of a 10 MHz spectrum allocation. The IFFT size can then, for example, be selected as $N = 1024$. This corresponds to a sampling rate $f_s = N \cdot \Delta f = 15.36$ MHz, where $\Delta f = 15$ kHz is the LTE subcarrier spacing.

[2] An OFDM signal defined according to (3.1) in theory has an infinite bandwidth and thus the sampling theorem can never be fulfilled completely.

[3] From now on the index m on the modulation symbols, indicating the OFDM symbol number, will be ignored unless especially needed.

Copyright Elsevier 2024

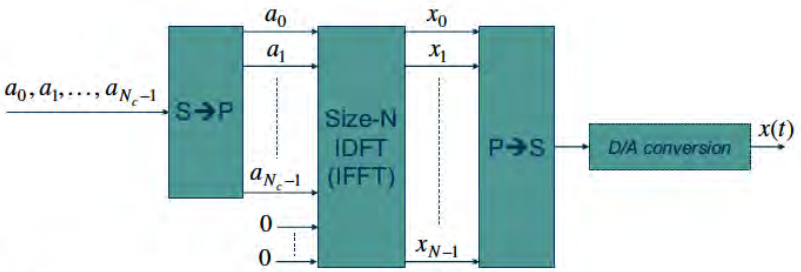


FIGURE 3.6

OFDM modulation by means of IFFT processing

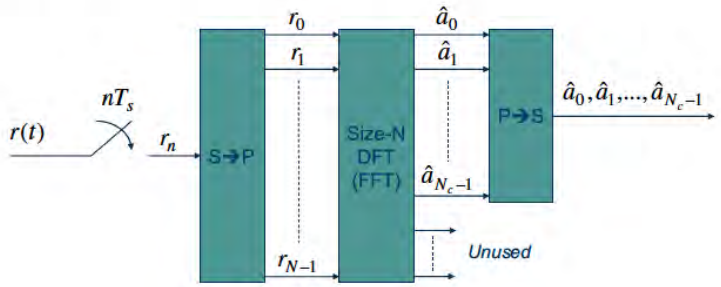


FIGURE 3.7

OFDM demodulation by means of FFT processing

It is important to understand that IDFT/IFFT-based implementation of an OFDM modulator, and even more so the exact IDFT/IFFT size, are just transmitter-implementation choices and not something that would be mandated by any radio-access specification. As an example, nothing forbids the implementation of an OFDM modulator as a set of parallel modulators, as illustrated in Figure 3.3. Also, nothing prevents the use of a larger IFFT size, for example a size-2048 IFFT, even in the case of a smaller number of OFDM subcarriers.

Similar to OFDM modulation, efficient FFT processing can be used for OFDM demodulation, replacing the bank of N_c parallel demodulators of Figure 3.5 with sampling rate $f_s = 1/T_s$, followed by a size-N DFT/FFT, as illustrated in Figure 3.7.

3.4 Cyclic-prefix insertion

As described in Section 3.2, an uncorrupted OFDM signal can be demodulated without any interference between subcarriers. One way to understand this *subcarrier orthogonality* is to

Copyright Elsevier 2024

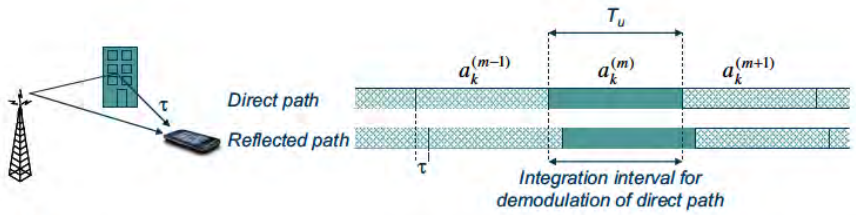


FIGURE 3.8

Time dispersion and corresponding received-signal timing

recognize that a modulated subcarrier $x_k(t)$ in (3.1) consists of an integer number of periods of complex exponentials during the demodulator integration interval $T_u = 1/\Delta f$.

However, in the case of a time-dispersive channel the orthogonality between the subcarriers will, at least partly, be lost. The reason for this loss of subcarrier orthogonality in the case of a time-dispersive channel is that the demodulator correlation interval for one path will overlap with the symbol boundary of a different path, as illustrated in Figure 3.8. Thus, the integration interval will not necessarily correspond to an integer number of periods of complex exponentials of that path as the modulation symbols a_k may differ between consecutive symbol intervals. As a consequence, in the case of a time-dispersive channel there will not only be inter-symbol interference within a subcarrier, but also interference between subcarriers.

Another way to explain the interference between subcarriers in the case of a time-dispersive channel is to have in mind that time dispersion on the radio channel is equivalent to a frequency-selective channel frequency response. As clarified in Section 3.2, orthogonality between OFDM subcarriers is not simply due to frequency-domain separation but due to the *specific* frequency-domain structure of each subcarrier. Even if the frequency-domain channel is constant over a bandwidth corresponding to the main lobe of an OFDM subcarrier and only the subcarrier side lobes are corrupted due to the radio-channel frequency selectivity, the orthogonality between subcarriers will be lost with inter-subcarrier interference as a consequence. Due to the relatively large side lobes of each OFDM subcarrier, a relatively limited amount of time dispersion or, equivalently, a relatively modest radio-channel frequency selectivity may cause substantial interference between subcarriers.

To deal with this problem and to make an OFDM signal truly insensitive to time dispersion on the radio channel, so-called *cyclic-prefix insertion* is typically used in OFDM transmission. As illustrated in Figure 3.9, cyclic-prefix insertion implies that the last part of the OFDM symbol is copied and inserted at the beginning of the OFDM symbol. Cyclic-prefix insertion thus increases the length of the OFDM symbol from T_u to $T_u + T_{CP}$, where T_{CP} is the length of the cyclic prefix, with a corresponding reduction in the OFDM symbol rate as a consequence. As illustrated in the lower part of Figure 3.9, if the correlation at the receiver side is still only carried out over a time interval $T_u = 1/\Delta f$, subcarrier orthogonality will then also be preserved in the case of a time-dispersive channel, as long as the span of the time dispersion is shorter than the cyclic-prefix length.

Copyright Elsevier 2024

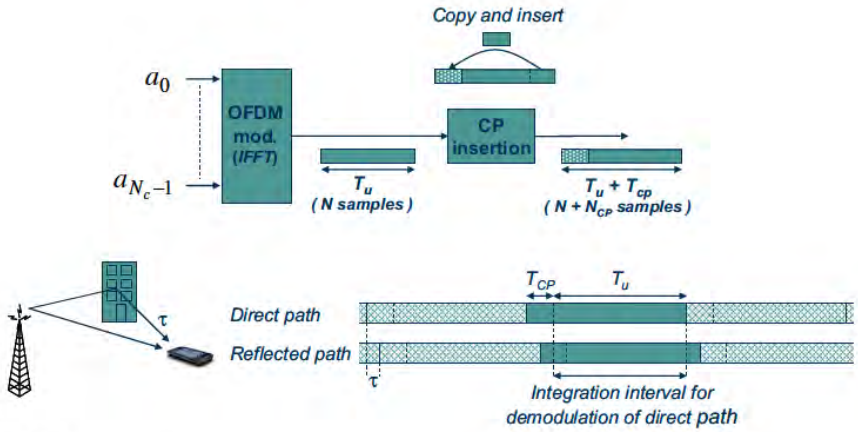


FIGURE 3.9

Cyclic-prefix insertion

In practice, cyclic-prefix insertion is carried out on the time-discrete output of the transmitter IFFT. Cyclic-prefix insertion then implies that the last N_{CP} samples of the IFFT output block of length N are copied and inserted at the beginning of the block, increasing the block length from N to $N + N_{CP}$. At the receiver side, the corresponding samples are discarded before OFDM demodulation by means of, for example, DFT/FFT processing.

Cyclic-prefix insertion is beneficial in the sense that it makes an OFDM signal insensitive to time dispersion as long as the span of the time dispersion does not exceed the length of the cyclic prefix. The drawback of cyclic-prefix insertion is that only a fraction $T_u/(T_u + T_{CP})$ of the received signal power is actually utilized by the OFDM demodulator, implying a corresponding power loss in the demodulation. In addition to this power loss, cyclic-prefix insertion also implies a corresponding loss in terms of bandwidth as the OFDM symbol rate is reduced without a corresponding reduction in the overall signal bandwidth.

One way to reduce the relative overhead due to cyclic-prefix insertion is to reduce the subcarrier spacing Δf, with a corresponding increase in the symbol time T_u as a consequence. However, this will increase the sensitivity of the OFDM transmission to fast channel variations—that is, high Doppler spread—as well as different types of frequency errors (see further Section 3.8).

It is also important to understand that the cyclic prefix does not necessarily have to cover the entire length of the channel time dispersion. In general, there is a trade-off between the power loss due to the cyclic prefix and the signal corruption (inter-symbol and inter-subcarrier interference) due to residual time dispersion not covered by the cyclic prefix. At a certain point, further reduction of the signal corruption due to further increase of the cyclic-prefix length will not justify the corresponding additional power loss. This also means that,

Copyright Elsevier 2024

although the amount of time dispersion typically increases with the cell size, beyond a certain cell size there is often no reason to increase the cyclic prefix further as the corresponding power loss due to a further increase of the cyclic prefix would have a larger negative impact, compared to the signal corruption due to the residual time dispersion not covered by the cyclic prefix [16].

3.5 Frequency-domain model of OFDM transmission

Assuming a sufficiently large cyclic prefix, the linear convolution of a time-dispersive radio channel will appear as a circular convolution during the demodulator integration interval T_u. The combination of OFDM modulation (IFFT processing), a time-dispersive radio channel, and OFDM demodulation (FFT processing) can then be seen as a *frequency-domain* channel as illustrated in Figure 3.10, where the frequency-domain channel taps $H_0, \ldots, H_{N_c-1}$ can be directly derived from the channel impulse response.

The demodulator output b_k in Figure 3.10 is the transmitted modulation symbol a_k scaled and phase rotated by the complex frequency-domain channel tap H_k and impaired by noise n_k. To properly recover the transmitted symbol for further processing, for example data demodulation and channel decoding, the receiver should multiply b_k by the complex conjugate of H_k, as illustrated in Figure 3.11. This is often expressed as a *one-tap equalizer* being applied to each received subcarrier.

3.6 Channel estimation and reference symbols

As described above, to demodulate the transmitted modulation symbol a_k and allow for proper decoding of the transmitted information at the receiver side, scaling with the complex

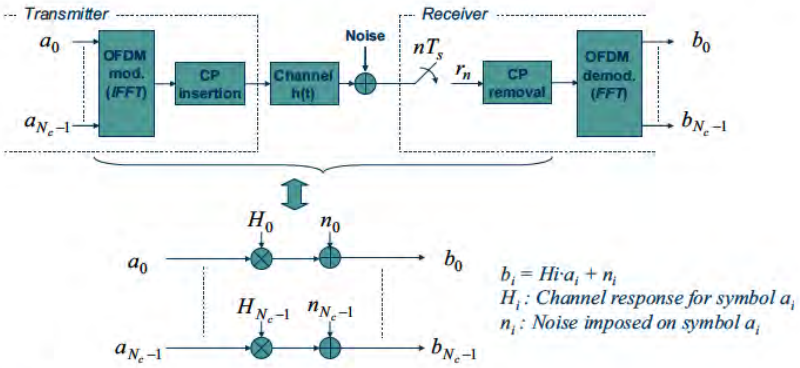


FIGURE 3.10

Frequency-domain model of OFDM transmission/reception

Copyright Elsevier 2024

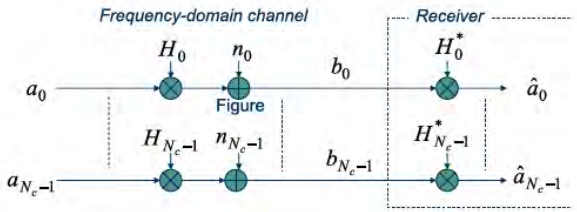


FIGURE 3.11

Frequency-domain model of OFDM transmission/reception with "one-tap equalization" at the receiver

conjugate of the frequency-domain channel tap H_k should be applied after OFDM demodulation (*FFT* processing) (see Figure 3.11). To be able to do this, the receiver needs an estimate of the frequency-domain channel taps $H_0, \ldots, H_{N_c-1}$.

The frequency-domain channel taps can be estimated indirectly by first estimating the channel impulse response and, from that, calculating an estimate of H_k. However, a more straightforward approach is to estimate the frequency-domain channel taps directly. This can be done by inserting known *reference symbols*, sometimes also referred to as *pilot symbols*, at regular intervals within the OFDM time–frequency grid, as illustrated in Figure 3.12. Using knowledge about the reference symbols, the receiver can estimate the frequency-domain channel around the location of the reference symbol. The reference symbols should also have a sufficiently high density in both the time and the frequency domains to be able to provide estimates for the entire time–frequency grid in the case of radio channels subject to high frequency and/or time selectivity.

Different more or less advanced algorithms can be used for the channel estimation, ranging from simple averaging in combination with linear interpolation to Minimum-Mean-Square-Error (MMSE) estimation relying on more detailed knowledge of the

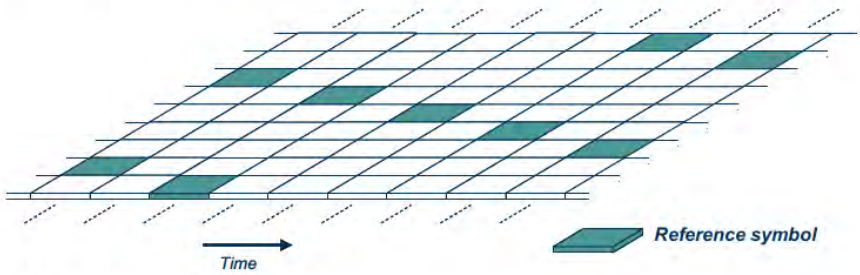


FIGURE 3.12

Time-frequency grid with known reference symbols

Copyright Elsevier 2024

channel time/frequency-domain characteristics. Readers are referred to, for example, [17] for a more in-depth discussion on channel estimation for OFDM.

3.7 Frequency diversity with OFDM: importance of channel coding

As discussed in Section 2.3 in the previous chapter, a radio channel is always subject to some degree of frequency selectivity, implying that the channel quality will vary in the frequency domain. In the case of a single wideband carrier, such as a WCDMA carrier, each modulation symbol is transmitted over the entire signal bandwidth. Thus, in the case of the transmission of a single wideband carrier over a highly frequency-selective channel (see Figure 3.13a), each modulation symbol will be transmitted both over frequency bands with relatively good quality (relatively high signal strength) and frequency bands with low quality (low signal strength). Such transmission of information over multiple frequency bands with different instantaneous channel quality is also referred to as *frequency diversity*.

On the other hand, in the case of OFDM transmission each modulation symbol is mainly confined to a relatively narrow bandwidth. Thus, for OFDM transmission over a frequency-selective channel, certain modulation symbols may be fully confined to a frequency band with very low instantaneous signal strength, as illustrated in Figure 3.13b. The individual modulation symbols will typically not experience any substantial frequency diversity even if the channel is highly frequency selective over the overall OFDM transmission bandwidth. As a consequence, the basic error-rate performance of OFDM transmission over a frequency-selective channel is relatively poor and especially much worse than the basic error rate in the case of a single wideband carrier.

However, in practice channel coding is used in most cases of digital communication and especially in mobile communication. Channel coding implies that each bit of information to be transmitted is spread over several, often very many, code bits. If these coded bits are then,

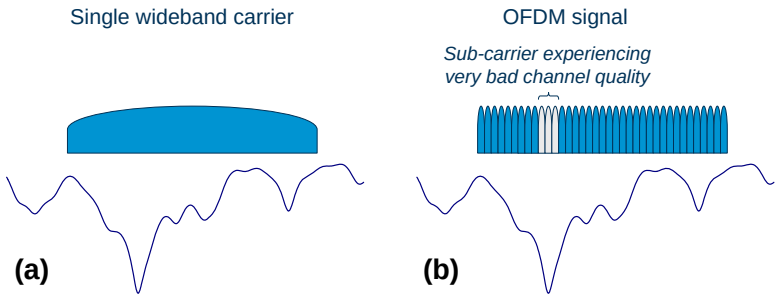


FIGURE 3.13

(a) Transmission of single wideband carrier and (b) OFDM transmission over a frequency-selective channel

Copyright Elsevier 2024

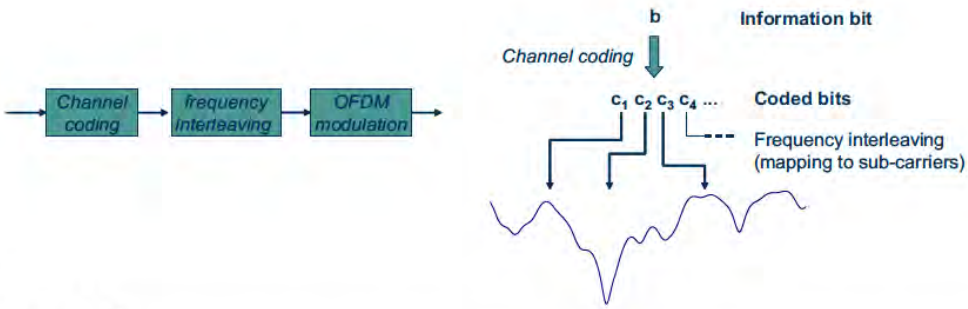


FIGURE 3.14

Channel coding in combination with frequency-domain interleaving to provide frequency diversity in case of OFDM transmission

via modulation symbols, mapped to a set of OFDM subcarriers that are well distributed over the overall transmission bandwidth of the OFDM signal, as illustrated in Figure 3.14, each information bit will experience frequency diversity in the case of transmission over a radio channel that is frequency selective over the transmission bandwidth, despite the fact that the subcarriers, and thus also the code bits, will not experience any frequency diversity. Distributing the code bits in the frequency domain, as illustrated in Figure 3.14, is sometimes referred to as *frequency interleaving*. This is similar to the use of time-domain interleaving to benefit from channel coding in the case of fading that varies in time.

In contrast to the transmission of a single wideband carrier, channel coding (combined with frequency interleaving) is an essential component in order for OFDM transmission to be able to benefit from frequency diversity on a frequency-selective channel. As channel coding is typically used in most cases of mobile communication this is not a very serious drawback, especially taking into account that a significant part of the available frequency diversity can be captured already with a relatively high code rate.

3.8 Selection of basic OFDM parameters

If OFDM is to be used as the transmission scheme in a mobile-communication system, the following basic OFDM parameters need to be decided upon:

- The subcarrier spacing Δf.
- The number of subcarriers N_c, which, together with the subcarrier spacing, determines the overall transmission bandwidth of the OFDM signal.
- The cyclic-prefix length T_{CP}. Together with the subcarrier spacing $\Delta f = 1/T_u$, the cyclic-prefix length determines the overall OFDM symbol time $T = T_{CP} + T_u$ or, equivalently, the OFDM symbol rate.

Copyright Elsevier 2024

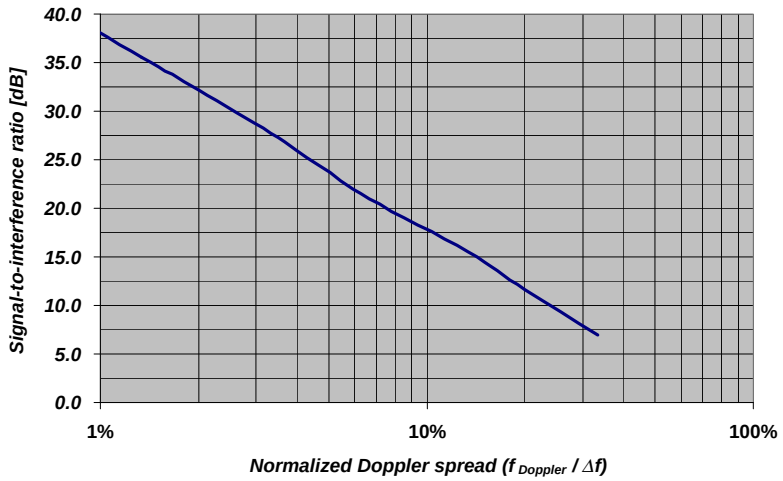


FIGURE 3.15

Subcarrier interference as a function of the normalized Doppler spread $f_{Doppler}/\Delta f$

3.8.1 OFDM subcarrier spacing

There are two factors that constrain the selection of the OFDM subcarrier spacing:

- The OFDM subcarrier spacing should be as small as possible (T_u as large as possible) to minimize the relative cyclic-prefix overhead $T_{CP}/(T_u + T_{CP})$—see further Section 3.8.3.
- A too small subcarrier spacing increases the sensitivity of the OFDM transmission to Doppler spread and different kinds of frequency inaccuracies.

A requirement for the OFDM subcarrier orthogonality (3.2) to hold at the receiver side—that is, after the transmitted signal has propagated over the radio channel—is that the instantaneous channel does not vary noticeably during the demodulator correlation interval T_u (see Figure 3.5). In the case of such channel variations, for example due to very high Doppler spread, the orthogonality between subcarriers will be lost, with inter-subcarrier interference as a consequence. Figure 3.15 illustrates the subcarrier signal-to-interference ratio due to inter-subcarrier interference between two neighboring subcarriers, as a function of the normalized Doppler spread. When considering Figure 3.15, it should be borne in mind that a subcarrier will be subject to interference from multiple subcarriers on both sides[4]—that is, the overall inter-subcarrier interference from all subcarriers will be higher than what is illustrated in Figure 3.15.

In practice, the amount of inter-subcarrier interference that can be accepted very much depends on the service to be provided and to what extent the received signal is corrupted due

[4]Except for the subcarrier at the edge of the spectrum.

Copyright Elsevier 2024

to noise and other impairments. As an example, on the cell border of large cells the signal-to-noise/interference ratio will be relatively low, with relatively low achievable data rates as a consequence. A small amount of additional inter-subcarrier interference, for example due to Doppler spread, may then be more or less negligible. At the same time, in high signal-to-noise/interference scenarios, for example in small cells with low traffic or close to the base station, where high data rates are to be provided, the same amount of inter-subcarrier interference may have a much more negative impact.

It should also be noted that, in addition to Doppler spread, inter-subcarrier interference will also be due to different transmitter and receiver inaccuracies, such as frequency errors and phase noise.

3.8.2 Number of subcarriers

Once the subcarrier spacing has been selected based on environment, expected Doppler spread and time dispersion, etc., the number of subcarriers can be determined based on the amount of spectrum available and the acceptable out-of-band emissions.

The basic bandwidth of an OFDM signal equals $N_c \cdot \Delta f$, that is, the number of subcarriers multiplied by the subcarrier spacing. However, as can be seen in Figure 3.16, the spectrum of a basic OFDM signal falls off very slowly outside the basic OFDM bandwidth and especially much slower than for a WCDMA signal. The reason for the large *out-of-band emission* of a basic OFDM signal is the use of rectangular pulse shaping (Figure 3.1), leading to per-subcarrier side lobes that fall off relatively slowly. However, in practice, straightforward filtering or *time-domain windowing* [18] will be used to suppress a main part of the OFDM

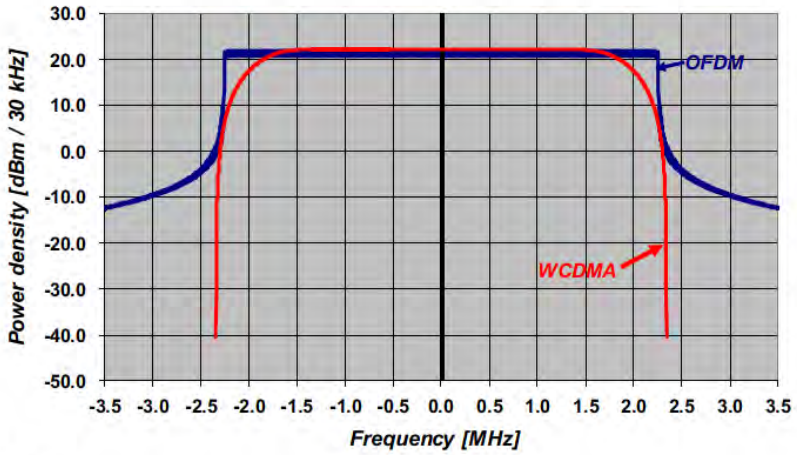


FIGURE 3.16

Spectrum of a basic 5 MHz OFDM signal compared with WCDMA spectrum

Copyright Elsevier 2024

out-of-band emissions. In practice, typically of the order of 10% guard-band is needed for an OFDM signal implying, as an example, that in a spectrum allocation of 5 MHz, the basic OFDM bandwidth $N_c \cdot \Delta f$ could be of the order of 4.5 MHz. Assuming, for example, a subcarrier spacing of 15 kHz as selected for LTE, this corresponds to approximately 300 subcarriers in 5 MHz.

3.8.3 Cyclic-prefix length

In principle, the cyclic-prefix length T_{CP} should cover the maximum length of the time dispersion expected to be experienced. However, as already discussed, increasing the length of the cyclic prefix, without a corresponding reduction in the subcarrier spacing Δf, implies an additional overhead in terms of power as well as bandwidth. In particular, the power loss implies that, as the cell size grows and the system performance becomes more power limited, there is a trade-off between the loss in power due to the cyclic prefix and the signal corruption due to time dispersion not covered by the cyclic prefix. As already mentioned, this implies that, although the amount of time dispersion typically increases with the cell size, beyond a certain cell size there is often no reason to increase the cyclic prefix further as the corresponding power loss would have a larger negative impact, compared to the signal corruption due to the residual time dispersion not covered by the cyclic prefix [16].

One situation where a longer cyclic prefix may be needed is in the case of multi-cell transmission using SFN (Single-Frequency Network), as further discussed in Section 3.11.

Thus, to be able to optimize performance to different environments, some OFDM-based systems support multiple cyclic-prefix lengths. The various cyclic-prefix lengths can then be used in different transmission scenarios:

- Shorter cyclic prefix in small-cell environments to minimize the cyclic-prefix overhead.
- Longer cyclic prefix in environments with extreme time dispersion and especially in the case of SFN operation.

3.9 Variations in instantaneous transmission power

According to Section 2.3.1, one of the drawbacks of multi-carrier transmission is the corresponding large variations in the instantaneous transmit power, implying a reduced power-amplifier efficiency and higher mobile-terminal power consumption; alternatively, that the power-amplifier output power has to be reduced with a reduced range as a consequence. Being a kind of multi-carrier transmission scheme, OFDM is subject to the same drawback.

However, a large number of different methods have been proposed to reduce the large power peaks of an OFDM signal:

- In the case of *tone reservation* [19], a subset of the OFDM subcarriers are not used for data transmission. Instead, these subcarriers are modulated in such a way that the largest

Copyright Elsevier 2024

peaks of the overall OFDM signal are suppressed, allowing for a reduced power-amplifier back-off. One drawback of tone reservation is the bandwidth loss due to the fact that a number of subcarriers are not available for actual data transmission. The calculation of what modulation to apply to the reserved tones can also be of relatively high complexity.

- In the case of *prefiltering* or *precoding*, linear processing is applied to the sequence of modulation symbols before OFDM modulation. DFTS-spread OFDM (DFTS-OFDM) as described in the next chapter, and which is used for the LTE uplink, can be seen as one kind of prefiltering.
- In the case of *selective scrambling* [20], the coded-bit sequence to be transmitted is scrambled with a number of different scrambling codes. Each scrambled sequence is then OFDM modulated and the signal with the lowest peak power is selected for transmission. After OFDM demodulation at the receiver side, descrambling and subsequent decoding is carried out for all the possible scrambling sequences. Only the decoding carried out for the scrambling code actually used for the transmission will provide a correct decoding result. A drawback of selective scrambling is increased receiver complexity as multiple decodings need to be carried out in parallel.

Readers are referred to the parenthetical references above for a more in-depth discussion on different peak-reduction schemes.

3.10 OFDM as a user-multiplexing and multiple-access scheme

The discussion has, until now, implicitly assumed that all OFDM subcarriers are transmitted from the same transmitter to a certain receiver, that is:

- downlink transmission of all subcarriers to a *single* terminal;
- uplink transmission of all subcarriers from a *single* terminal.

However, OFDM can also be used as a *user-multiplexing* or *multiple-access scheme*, allowing for simultaneous frequency-separated transmissions to/from multiple terminals (see Figure 3.17).

In the downlink direction, OFDM as a user-multiplexing scheme implies that, in each OFDM symbol interval, different subsets of the overall set of available subcarriers are used for transmission to *different* terminals (see Figure 3.17a).

Similarly, in the uplink direction, OFDM as a *user-multiplexing* or *multiple-access scheme* implies that, in each OFDM symbol interval, different subsets of the overall set of subcarriers are used for data transmission from *different* terminals (see Figure 3.17b). In this case, the term Orthogonal Frequency-Division Multiple Access or OFDMA is also often used.[5]

[5]The term OFDMA is sometimes also used to denote the use of OFDM to multiplex multiple users in the downlink, as illustrated in Figure 3.17a.

Copyright Elsevier 2024

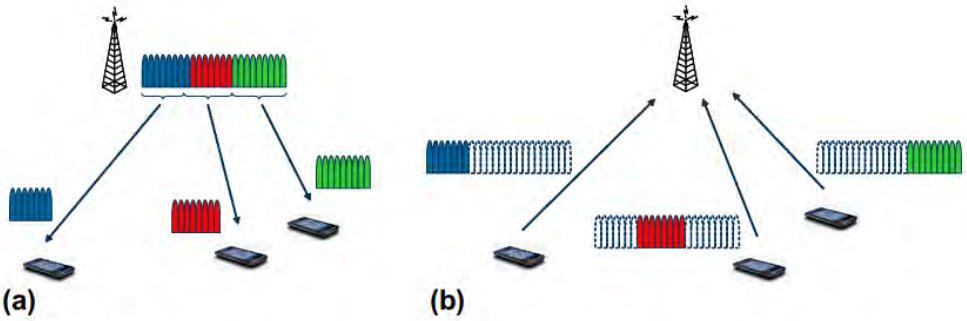


FIGURE 3.17

OFDM as a user-multiplexing/multiple-access scheme: (a) downlink and (b) uplink

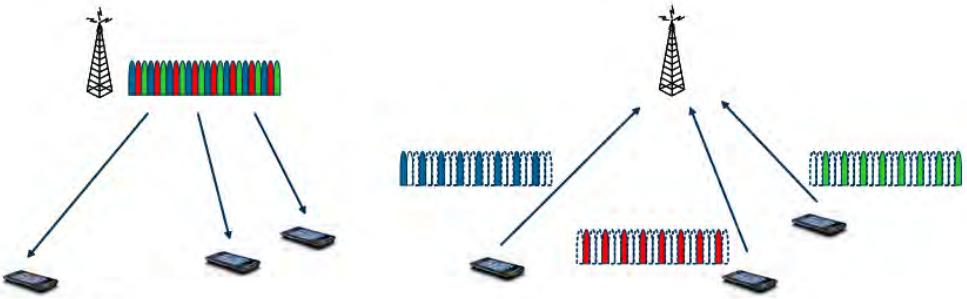


FIGURE 3.18

Distributed user multiplexing

Figure 3.17 assumes that *consecutive* subcarriers are used for transmission to/from the same terminal. However, distributing the subcarriers to/from a terminal in the frequency domain is also possible, as illustrated in Figure 3.18. The benefit of such *distributed* user multiplexing or *distributed* multiple access is a possibility for additional frequency diversity as each transmission is spread over a wider bandwidth.

In the case when OFDMA is used as an uplink multiple-access scheme—that is, in the case of frequency multiplexing of OFDM signals from multiple terminals—it is critical that the transmissions from the different terminals arrive approximately time aligned at the base station. More specifically, the transmissions from the different terminals should arrive at the base station with a timing misalignment less than the length of the cyclic prefix to preserve orthogonality between subcarriers received from different terminals and thus avoid inter-user interference.

Due to the differences in distance to the base station for different terminals and the corresponding differences in the propagation time (which may far exceed the length of the cyclic

Copyright Elsevier 2024

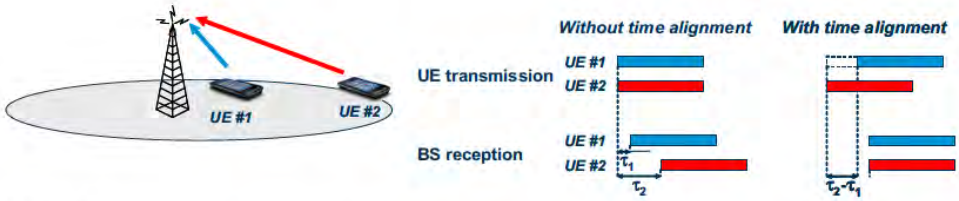


FIGURE 3.19

Uplink transmission-timing control

prefix), it is therefore necessary to control the uplink transmission timing of each terminal (see Figure 3.19). Such *transmission-timing control* should adjust the transmit timing of each terminal to ensure that uplink transmissions arrive approximately time aligned at the base station. As the propagation time changes as the terminal is moving within the cell, the transmission-timing control should be an active process, continuously adjusting the exact transmit timing of each terminal.

Furthermore, even in the case of perfect transmission-timing control, there will always be some interference between subcarriers, for example due to frequency errors. Typically this interference is relatively low in the case of reasonable frequency errors, Doppler spread, etc. (see Section 3.8). However, this assumes that the different subcarriers are received with at least approximately the same power. In the uplink, the propagation distance and thus the path loss of the different mobile-terminal transmissions may differ significantly. If two terminals are transmitting with the same power, the received-signal strengths may differ significantly, implying a potentially significant interference from the stronger signal to the weaker signal unless the subcarrier orthogonality is perfectly retained. To avoid this, at least some degree of *uplink transmission-power control* may need to be applied in case of uplink OFDMA, reducing the transmission power of user terminals close to the base station and ensuring that all received signals will be of approximately the same power.

3.11 Multi-cell broadcast/multicast transmission and OFDM

The provision of broadcast/multicast services in a mobile-communication system implies that the same information is to be *simultaneously* provided to multiple terminals, often dispersed over a large area corresponding to a large number of cells, as shown in Figure 3.20. The broadcast/multicast information may be a TV news clip, information about the local weather conditions, stock-market information, or any other kind of information that, at a given time, may be of interest to a large number of people.

When the same information is to be provided to multiple terminals within a cell it is often beneficial to provide this information as a single "broadcast" radio transmission covering the entire cell and simultaneously being received by all relevant terminals (Figure 3.21a), rather than providing the information by means of individual transmissions to each terminal (unicast transmission; see Figure 3.21b).

Copyright Elsevier 2024

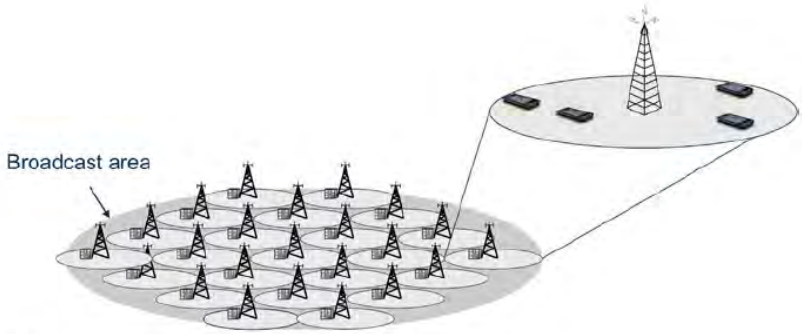


FIGURE 3.20

Broadcast scenario

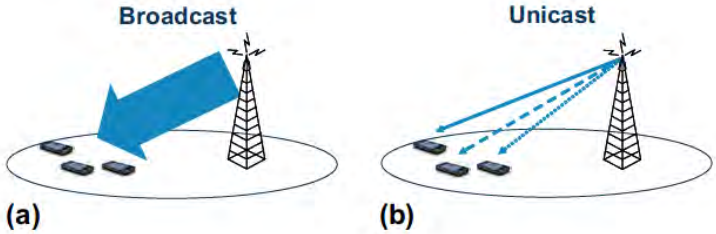


FIGURE 3.21

Broadcast vs. unicast transmission: (a) broadcast and (b) unicast

As a broadcast transmission according to Figure 3.21 has to be dimensioned to reach the worst-case terminals, including terminals at the cell border, it will be relative costly in terms of the recourses (base-station transmit power) needed to provide a certain broadcast-service data rate. Alternatively, taking into account the limited signal-to-noise ratio that can be achieved at, for example, the cell edge, the achievable broadcast data rates may be relatively limited, especially in the case of large cells. One way to increase the broadcast data rates would then be to reduce the cell size, thereby increasing the cell-edge receive power. However, this will increase the number of cells to cover a certain area and is thus obviously negative from a cost-of-deployment point of view.

However, as discussed above, the provision of broadcast/multicast services in a mobile-communication network typically implies that identical information is to be provided over a large number of cells. In such a case, the resources (downlink transmit power) needed to provide a certain broadcast data rate can be considerably reduced if terminals at the cell edge can utilize the received power from broadcast transmissions from multiple cells when detecting/decoding the broadcast data.

Copyright Elsevier 2024

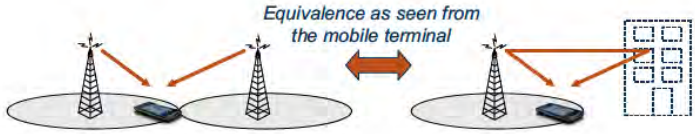


FIGURE 3.22

Equivalence between simulcast transmission and multi-path propagation

One way to mitigate this and further improve the provision of broadcast/multicast services in a mobile-communication network is to ensure that the broadcast transmissions from different cells *are truly identical* and *transmitted mutually time aligned*. In this case, the transmissions received from multiple cells will, as seen from the terminal, appear as a single transmission subject to severe multi-path propagation, as illustrated in Figure 3.22. The transmission of identical time-aligned signals from multiple cells, especially in the case of broadcast/multicast services, is sometimes referred to as Single-Frequency Network (SFN) operation [21].

In the case of such identical time-aligned transmissions from multiple cells, the "inter-cell interference" due to transmissions in neighboring cells will, from a terminal point of view, be replaced by signal corruption due to time dispersion. If the broadcast transmission is based on OFDM with a cyclic prefix that covers the main part of this "time dispersion," the achievable broadcast data rates are thus only limited by noise, implying that, especially in smaller cells, very high broadcast data rates can be achieved. Furthermore, the OFDM receiver does not need to explicitly identify the cells to be soft combined. Rather, all transmissions that fall within the cyclic prefix will "automatically" be captured by the receiver.

Copyright Elsevier 2024

CHAPTER

Wider-Band "Single-Carrier" Transmission

4

The previous chapters discussed multi-carrier transmission in general (Section 2.3.1) and OFDM transmission in particular (Chapter 3) as means to allow for very high overall transmission bandwidth while still being robust enough to signal corruption due to radio-channel frequency selectivity.

However, as discussed in the previous chapter, a drawback of OFDM modulation, as well as any kind of multi-carrier transmission, is the large variations in the instantaneous power of the transmitted signal. Such power variations imply a reduced power-amplifier efficiency and higher power-amplifier cost. This is especially critical for the uplink, due to the high importance of low mobile-terminal power consumption and cost.

As discussed in Chapter 3, several methods have been proposed on how to reduce the large power variations of an OFDM signal. However, most of these methods have limitations as to what extent the power variations can be reduced. Furthermore, most of the methods also imply a significant computational complexity and/or a reduced link performance.

Thus, there is interest to consider wider-band *single-carrier* transmission as an alternative to multi-carrier transmission, especially for the uplink—that is, for mobile-terminal transmissions. It is then necessary to consider what can be done to handle the corruption to the signal waveform that will occur in most mobile-communication environments due to radio-channel frequency selectivity.

4.1 Equalization against radio-channel frequency selectivity

Historically, the main method to handle signal corruption due to radio-channel frequency selectivity has been to apply different forms of *equalization* [12] at the receiver side. The aim of equalization is to, by different means, compensate for the channel frequency selectivity and thus, at least to some extent, restore the original signal shape.

4.1.1 Time-domain linear equalization

The most basic approach to equalization is the time-domain linear equalizer, consisting of a linear filter with an impulse response $w(\tau)$ applied to the received signal (see Figure 4.1).

4G: LTE/LTE-Advanced for Mobile Broadband. http://dx.doi.org/10.1016/B978-0-12-419985-9.00004-0
Copyright © 2014 Erik Dahlman, Stefan Parkvall and Johan Sköld. Published by Elsevier Ltd. All rights reserved.

Copyright Elsevier 2024

Samsung Exhibit 1017, Page 75 of 538
Samsung Electronics Co., Ltd. v. Asus Technology Licensing Inc.
IPR2024-01217

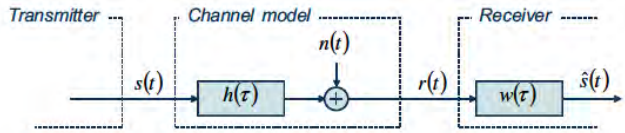


FIGURE 4.1

General time-domain linear equalization

By selecting different filter impulse responses, different receiver/equalizer strategies can be implemented. As an example, in DS-CDMA-based systems a so-called RAKE receiver structure has historically often been used. The RAKE receiver is simply the receiver structure of Figure 4.1 where the filter impulse response has been selected to provide *channel-matched filtering*:

$$w(\tau) = h^*(-\tau) \tag{4.1}$$

That is, the filter response has been selected as the complex conjugate of the time-reversed channel impulse response. This is also often referred to as a *Maximum-Ratio Combining* (MRC) filter setting [12].

Selecting the receiver filter according to the MRC criterion—that is, as a channel-matched filter—maximizes the post-filter signal-to-noise ratio (thus the term maximum-ratio combining). However, MRC-based filtering does not provide any compensation for any radio-channel frequency selectivity—that is, no equalization. Thus, MRC-based receiver filtering is appropriate when the received signal is mainly impaired by noise or interference from other transmissions but not when a main part of the overall signal corruption is due to the radio-channel frequency selectivity.

Another alternative is to select the receiver filter to fully compensate for the radio-channel frequency selectivity. This can be achieved by selecting the receiver-filter impulse response to fulfill the relation:

$$h(\tau) \otimes w(\tau) = 1 \tag{4.2}$$

where "$\otimes$" denotes linear convolution. This selecting of the filter setting, also known as *Zero-Forcing* (ZF) equalization [12], provides full compensation for any radio-channel frequency selectivity (complete equalization) and thus full suppression of any related signal corruption. However, zero-forcing equalization may lead to a large, potentially very large, increase in the noise level after equalization and to an overall degradation in the link performance. This will be the case especially when the channel has large variations in its frequency response.

A third and, in most cases, better alternative is to select a filter setting that provides a trade-off between signal corruption due to radio-channel frequency selectivity and noise/interference. This can, for example, be done by selecting the filter to minimize the mean-square error between the equalizer output and the transmitted signal—that is, to minimize:

$$\varepsilon = E\left\{\left|\hat{s}(t) - s(t)\right|^2\right\}. \tag{4.3}$$

Copyright Elsevier 2024

Samsung Exhibit 1017, Page 76 of 538
Samsung Electronics Co., Ltd. v. Asus Technology Licensing Inc.
IPR2024-01217

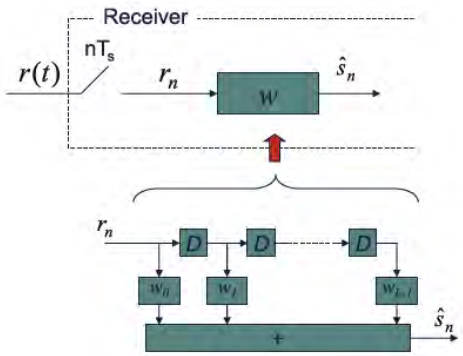


FIGURE 4.2

Linear equalization implemented as a time-discrete FIR filter

This is also referred to as a *Minimum Mean-Square Error* (MMSE) equalizer setting [12].

In practice, the linear equalizer has most often been implemented as a *time-discrete FIR filter* [22] with L filter taps applied to the sampled received signal, as illustrated in Figure 4.2. In general, the complexity of such a time-discrete equalizer grows relatively rapidly with the bandwidth of the signal to be equalized:

- A more wideband signal is subject to relatively more radio-channel frequency selectivity or, equivalently, relatively more time dispersion. This implies that the equalizer needs to have a larger span (larger length L—that is, more filter taps) to be able to properly compensate for the channel frequency selectivity.
- A more wideband signal leads to a correspondingly higher sampling rate for the received signal. Thus, the receiver-filter processing needs to be carried out with a correspondingly higher rate.

It can be shown [23] that the time-discrete MMSE equalizer setting $\overline{w} = [w_0, w_1, \ldots, w_{L-1}]^H$ is given by the expression:

$$\overline{w} = R^{-1}\overline{p}. \tag{4.4}$$

In this expression, R is the *channel-output auto-correlation matrix* of size $L \times L$, which depends on the channel impulse response as well as on the noise level, and $\overline{p}$ is the *channel-output/channel-input cross-correlation vector* of size $L \times 1$ that depends on the channel impulse response.

In particular, in the case of a large equalizer span (large L), the time-domain MMSE equalizer may be of relatively high complexity:

- The equalization itself (the actual filtering) may be of relatively high complexity according to the above.
- Calculation of the MMSE equalizer setting, especially the calculation of the inverse of the size $L \times L$ correlation matrix R, may be of relatively high complexity.

Copyright Elsevier 2024

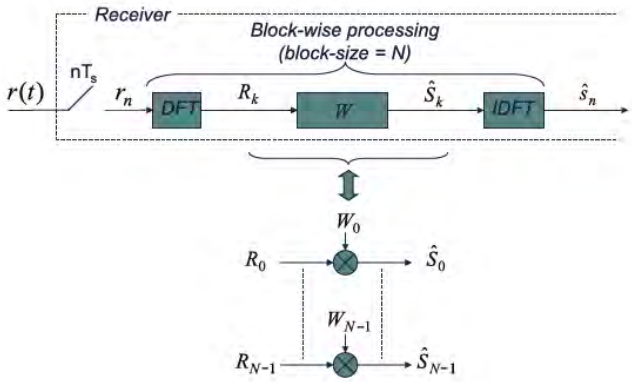


FIGURE 4.3

Frequency-domain linear equalization

4.1.2 Frequency-domain equalization

A possible way to reduce the complexity of linear equalization is to carry out the equalization in the frequency domain [24], as illustrated in Figure 4.3. In such *frequency-domain linear equalization*, the equalization is carried out block-wise with block size-N. The sampled received signal is first transformed into the frequency domain by means of a size-N DFT. The equalization is then carried out as frequency-domain filtering, with the frequency-domain filter taps $W_0, \ldots, W_{N-1}$, for example, being the DFT of the corresponding time-domain filter taps $w_0, \ldots, w_{L-1}$ of Figure 4.2. Finally, the equalized frequency-domain signal is transformed back to the time domain by means of a size-N inverse DFT. The block size-N should preferably be selected as $N = 2^n$ for some integer n to allow for computational-efficient radix-2 FFT/IFFT implementation of the DFT/IDFT processing.

For each processing block of size-N, the frequency-domain equalization basically consists of:

- A size-N DFT/FFT.
- N complex multiplications (the frequency-domain filter).
- A size-N inverse DFT/FFT.

Especially in the case of channels with extensive frequency selectivity, implying the need for a large span of a time-domain equalizer (large equalizer length L), equalization in the frequency domain according to Figure 4.3 can be of significantly less complexity, compared to the time-domain equalization illustrated in Figure 4.2.

However, there are two issues with frequency-domain equalization:

- The time-domain filtering of Figure 4.2 implements a time-discrete *linear convolution.* In contrast, frequency-domain filtering according to Figure 4.3 corresponds to *circular*

Copyright Elsevier 2024

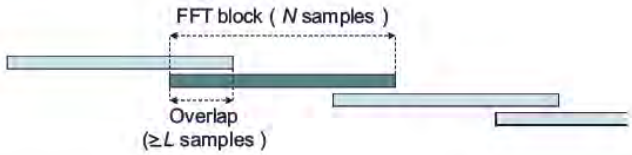


FIGURE 4.4

Overlap-and-discard processing

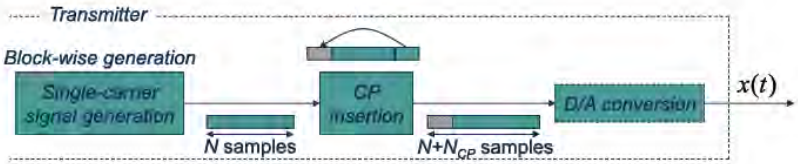


FIGURE 4.5

Cyclic-prefix insertion in the case of single-carrier transmission

convolution in the time domain. Assuming a time-domain equalizer of length L, this implies that the first $L-1$ samples at the output of the frequency-domain equalizer *will not* be identical to the corresponding output of the time-domain equalizer.

- The frequency-domain filter taps $W_0, \ldots, W_{N-1}$ can be determined by first determining the pulse response of the corresponding time-domain filter and then transforming this filter into the frequency domain by means of a DFT. However, as mentioned above, determining the MMSE time-domain filter may be relatively complex in the case of a large equalizer length L.

One way to address the first issue is to apply an *overlap* in the block-wise processing of the frequency-domain equalizer as outlined in Figure 4.4, where the overlap should be at least $L-1$ samples. With such an overlap, the first $L-1$ "incorrect" samples at the output of the frequency-domain equalizer can be *discarded* as the corresponding samples are also (correctly) provided as the last part of the previously received/equalized block. The drawback with this kind of "overlap-and-discard" processing is a computational overhead—that is, somewhat higher receiver complexity.

An alternative approach that addresses both of the above issues is to apply *cyclic-prefix insertion* at the transmitter side (see Figure 4.5). Similar to OFDM, cyclic-prefix insertion in the case of single-carrier transmission implies that a cyclic prefix of length N_{CP} samples is inserted block-wise at the transmitter side. The transmitter-side block size should be the same as the block size-N used for the receiver-side frequency-domain equalization.

With the introduction of a cyclic prefix, the channel will, from a receiver point of view, appear as a circular convolution over a receiver processing block of size-N. There is

Copyright Elsevier 2024

no need for any receiver overlap-and-discard processing. Furthermore, the frequency-domain filter taps can now be calculated directly from an estimate of the sampled channel frequency response without first determining the time-domain equalizer setting. As an example, in the case of an MMSE equalizer the frequency-domain filter taps can be calculated according to:

$$W_k = \frac{H_k^*}{|H_k|^2 + N_0} \tag{4.5}$$

where N_0 is the noise power and H_k is the sampled channel frequency response. For large equalizer lengths, this calculation is of much lower complexity compared to the time-domain calculation discussed in the previous section.

The drawback of cyclic-prefix insertion in the case of single-carrier transmission is the same as for OFDM—that is, it implies an overhead in terms of both power and bandwidth. One method to reduce the relative cyclic-prefix overhead is to increase the block size-N of the frequency-domain equalizer. However, for the block-wise equalization to be accurate, the channel needs to be approximately constant over a time span corresponding to the size of the processing block. This constraint provides an upper limit on the block size-N that depends on the rate of the channel variations. Note that this is similar to the constraint on the OFDM subcarrier spacing $\Delta f = 1/T_u$ depending on the rate of the channel variations, as discussed in Chapter 3.

4.1.3 Other equalizer strategies

The previous sections discussed different approaches to linear equalization as a means to counteract signal corruption of a wideband signal due to radio-channel frequency selectivity. However, there are also other approaches to equalization:

- *Decision-Feedback Equalization* (DFE) [12] implies that previously detected symbols are fed back and used to cancel the contribution of the corresponding transmitted symbols to the overall signal corruption. Such decision feedback is typically used in combination with time-domain linear filtering, where the linear filter transforms the channel response to a shape that is more suitable for the decision-feedback stage. Decision feedback can also very well be used in combination with frequency-domain linear equalization [24].
- *Maximum-Likelihood* (ML) detection, also known as *Maximum-Likelihood Sequence Estimation* (MLSE) [25], is strictly speaking not an equalization scheme but rather a receiver approach where the impact of radio-channel time dispersion is explicitly taken into account in the receiver-side detection process. Fundamentally, an ML detector uses the entire received signal to decide on the most likely transmitted sequence, taking into account the impact of the time dispersion on the received signal. To implement maximum-likelihood detection the *Viterbi algorithm* [26] is often used.[1] However, although maximum-likelihood detection based on the Viterbi algorithm has been

[1]The term *Viterbi equalizer* is sometimes used for the ML detector.

Copyright Elsevier 2024

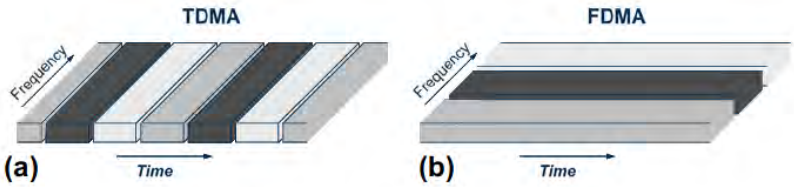


FIGURE 4.6

Orthogonal multiple access: (a) TDMA and (b) FDMA

extensively used for 2G mobile communication such as GSM, it can be too complex when applied to much wider transmission bandwidths, leading to both much more extensive channel frequency selectivity and much higher sampling rates.

4.2 Uplink FDMA with flexible bandwidth assignment

In practice, there are often multiple terminals within a cell and thus multiple terminals that should share the overall uplink radio resource within the cell by means of the *uplink intra-cell multiple-access scheme*.

In mobile communication based on WCDMA and CDMA2000, uplink transmissions within a cell are mutually *non-orthogonal* and the base-station receiver relies on the processing gain due to channel coding and additional direct-sequence spreading to suppress the intra-cell interference. Although non-orthogonal multiple access can, fundamentally, provide higher capacity compared to orthogonal multiple access, in practice the possibility for mutually orthogonal uplink transmissions within a cell is often beneficial from a system performance point of view.

Mutually orthogonal uplink transmissions within a cell can be achieved in the time domain (*Time-Division Multiple Access*, TDMA), as illustrated in Figure 4.6a. TDMA implies that different terminals within a cell transmit in different non-overlapping time intervals. In each such time interval, the full system bandwidth is then assigned for uplink transmission from a single terminal.

Alternatively, mutually orthogonal uplink transmissions within a cell can be achieved in the frequency domain (*Frequency-Division Multiple Access*, FDMA), as illustrated in Figure 4.6b—that is, by having terminals transmit in different frequency bands.

To be able to provide high-rate packet-data transmission, it should be possible to assign the entire system bandwidth for transmission from a single terminal. At the same time, due to the burstiness of most packet-data services, in many cases terminals will have no uplink data to transmit. Thus, for efficient packet-data access, a TDMA component should always be part of the uplink multiple-access scheme.

However, relying on only TDMA to provide orthogonality between uplink transmissions within a cell could be bandwidth inefficient, especially in the case of a very wide overall system

Copyright Elsevier 2024

Samsung Exhibit 1017, Page 81 of 538
Samsung Electronics Co., Ltd. v. Asus Technology Licensing Inc.
IPR2024-01217

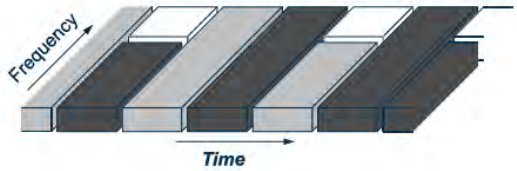


FIGURE 4.7

FDMA with flexible bandwidth assignment

bandwidth. As discussed in Chapter 2, a wide bandwidth is needed to support high data rates in a power-efficient way. However, the data rates that can be achieved over a radio link are, in many cases, limited by the available signal power (power-limited operation) rather than by the available bandwidth. This is especially the case for the uplink, due to the generally more limited mobile-terminal transmit power. Allocating the entire system bandwidth to a single terminal could, in such cases, be highly inefficient in terms of bandwidth utilization. As an example, allocating 20 MHz of transmission bandwidth to a terminal in a scenario where the achievable uplink data rate, due to mobile-terminal transmit-power limitations, is limited to a few 100 kbit/s would obviously imply a very inefficient usage of the overall available bandwidth. In such cases, a smaller transmission bandwidth should be assigned to the terminal and the remaining part of the overall system bandwidth should be used for uplink transmissions from other terminals. In addition to TDMA, an uplink transmission scheme should preferably allow for orthogonal user multiplexing also in the frequency domain—that is, FDMA.

At the same time, it should be possible to allocate the entire overall transmission bandwidth to a single terminal when the channel conditions are such that the wide bandwidth can be efficiently utilized—that is, when the achievable data rates are not power limited. Thus, an orthogonal uplink transmission scheme should allow for *FDMA with flexible bandwidth assignment*, as illustrated in Figure 4.7.

Flexible bandwidth assignment is straightforward to achieve with an OFDM-based uplink transmission scheme by dynamically allocating different numbers of subcarriers to different terminals depending on their instantaneous channel conditions. In the next section, it will be discussed how this can also be achieved in the case of low-PAPR "single-carrier" transmission, more specifically by means of so-called *DFT-spread OFDM*.

4.3 DFT-spread OFDM

DFT-spread OFDM (DFTS-OFDM) is a transmission scheme that can combine the desired properties discussed in the previous sections:

- Small variations in the instantaneous power of the transmitted signal ("single-carrier" property).
- Possibility for low-complexity high-quality equalization in the frequency domain.
- Possibility for FDMA with flexible bandwidth assignment.

Copyright Elsevier 2024

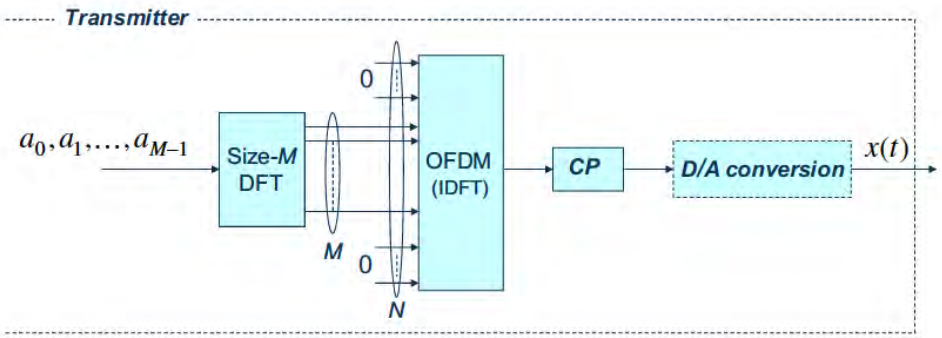


FIGURE 4.8

DFTS-OFDM signal generation

Due to these properties, DFTS-OFDM is, for example, used for uplink data transmission in LTE.

4.3.1 Basic principles

The basic principle of DFTS-OFDM transmission is illustrated in Figure 4.8. One way to interpret DFTS-OFDM is as normal OFDM with a *DFT-based precoding*. Similar to OFDM modulation, DFTS-OFDM relies on block-based signal generation. In the case of DFTS-OFDM, a block of M modulation symbols from some modulation alphabet, for example QPSK or 16QAM, is first applied to a size-M DFT. The output of the DFT is then applied to consecutive inputs (subcarriers) of an OFDM modulator where, in practice, the OFDM modulator will be implemented as a size-N *inverse* DFT (IDFT) with $N > M$ and where the unused inputs of the IDFT are set to zero. Typically, the IDFT size-N is selected as $N = 2^n$ for some integer n to allow for the IDFT to be implemented by means of computationally efficient radix-2 IFFT processing. Also, similar to normal OFDM, a cyclic prefix is preferably inserted for each transmitted block. As discussed in Section 4.1.2, the presence of a cyclic prefix allows for straightforward low-complexity frequency-domain equalization at the receiver side.

If the DFT size-M equals the IDFT size-N, the cascaded DFT/IDFT processing would completely cancel each other out. However, if M is smaller than N and the remaining inputs to the IDFT are set to zero, the output of the IDFT will be a signal with "single-carrier" properties—that is, a signal with low power variations, and with a bandwidth that depends on M. More specifically, assuming a sampling rate f_s at the output of the IDFT, the nominal bandwidth of the transmitted signal will be $BW = M/N \cdot f_s$. Thus, by varying the block size-M the instantaneous bandwidth of the transmitted signal can be varied, allowing for flexible-bandwidth assignment. Furthermore, by shifting the IDFT inputs to which the DFT outputs are mapped, the transmitted signal can be shifted in the frequency domain, as will be further discussed in Section 4.3.3.

Copyright Elsevier 2024

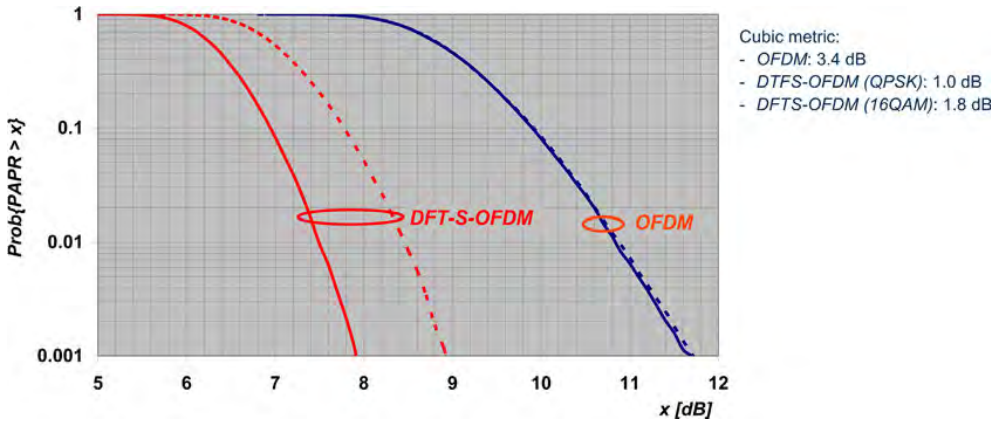


FIGURE 4.9

PAPR distribution for DFTS-OFDM and OFDM, respectively. Solid curve: QPSK. Dashed curve: 16QAM

To have a high degree of flexibility in the instantaneous bandwidth, given by the DFT size-M, it is typically not possible to ensure that M can be expressed as 2^m for some integer m. However, as long as M can be expressed as a product of relatively small prime numbers, the DFT can still be implemented as relatively low-complexity non-radix-2 FFT processing. As an example, a DFT size-$M = 144$ can be implemented by means of a combination of radix-2 and radix-3 FFT processing ($144 = 3^2 \cdot 2^4$).

The main benefit of DFTS-OFDM, compared to normal OFDM, is reduced variations in the instantaneous transmission power, implying the possibility for increased power-amplifier efficiency. This benefit of DFTS-OFDM is shown in Figure 4.9, which illustrates the distribution of the *Peak-to-Average-power Ratio* (PAPR) for DFTS-OFDM and conventional OFDM. The PAPR is defined as the peak power within one DFT block (one OFDM symbol) normalized by the average signal power. It should be noted that the PAPR distribution is *not* the same as the distribution of the instantaneous transmission power illustrated in Figure 2.3. Historically, PAPR distributions have often been used to illustrate the power variations of OFDM.

As can be seen in Figure 4.9, the PAPR is significantly lower for DFTS-OFDM, compared to OFDM. In the case of 16QAM modulation, the PAPR of DFTS-OFDM increases somewhat as expected (compare Figure 2.3). On the other hand, in the case of OFDM the PAPR distribution is more or less independent of the modulation scheme. The reason is that, as the transmitted OFDM signal is the sum of a large number of independently modulated subcarriers, the instantaneous power has an approximately exponential distribution, regardless of the modulation scheme applied to the different subcarriers.

Although the PAPR distribution can be used to qualitatively illustrate the difference in power variations between different transmission schemes, it is not a very good measure to more accurately quantify the impact of the power variations on, for example, the required power-amplifier back-off.

Copyright Elsevier 2024

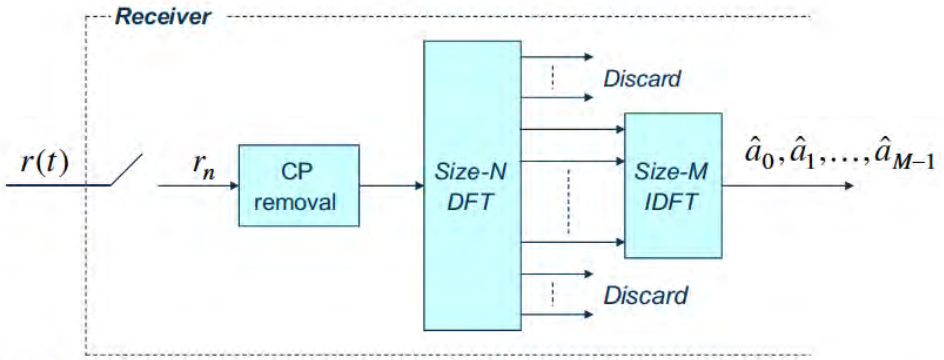


FIGURE 4.10

Basic principle of DFTS-OFDM demodulation

A better measure of the impact on the required power-amplifier back-off and the corresponding impact on the power-amplifier efficiency is given by the so-called *cubic metric* [27]. The cubic metric is a measure of the amount of additional back-off needed for a certain signal waveform, relative to the back-off needed for some reference waveform.

As can be seen in Figure 4.9, the cubic metric (given to the right of the graph) follows the same trend as the PAPR. However, the differences in the cubic metric are somewhat smaller than the corresponding differences in the PAPR.

4.3.2 DFTS-OFDM receiver

The basic principle of demodulation of a DFTS-OFDM signal is illustrated in Figure 4.10. The operations are basically the reverse of those for the DFTS-OFDM signal generation of Figure 4.8—that is, size-N DFT (FFT) processing, removal of the frequency samples not corresponding to the signal to be received, and size-M inverse DFT processing.

In the ideal case, with no signal corruption on the radio channel, DFTS-OFDM demodulation according to Figure 4.10 will perfectly restore the block of transmitted symbols. However, in the case of a time-dispersive or, equivalently, a frequency-selective radio channel, the DFTS-OFDM signal will be corrupted, with "self-interference" as a consequence. This can be understood in two ways:

- Being a wideband single-carrier signal, the DFTS-OFDM spread signal is corrupted in the case of a time-dispersive channel.
- If the channel is frequency selective over the span of the DFT, the inverse DFT at the receiver will not be able to correctly reconstruct the original block of transmitted symbols.

Thus, in the case of DFTS-OFDM, an equalizer is needed to compensate for the radio-channel frequency selectivity. Assuming the basic DFTS-OFDM demodulator structure

Copyright Elsevier 2024

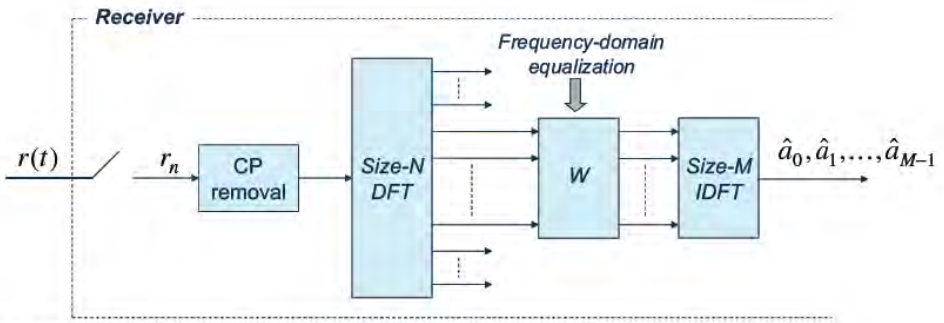


FIGURE 4.11

DFTS-OFDM demodulator with frequency-domain equalization

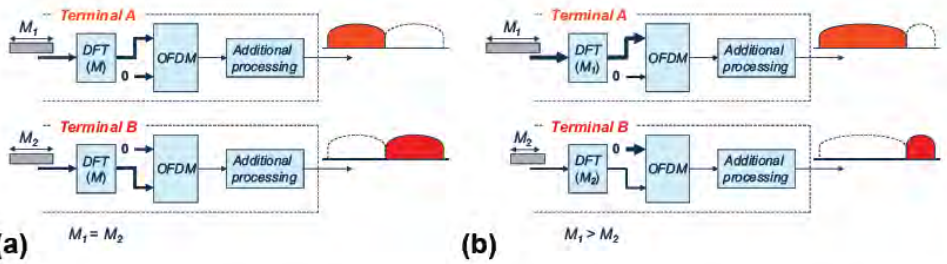


FIGURE 4.12

Uplink user multiplexing in the case of DFTS-OFDM. (a) Equal-bandwidth assignment and (b) unequal-bandwidth assignment

according to Figure 4.10, frequency-domain equalization as discussed in Section 4.1.2 is especially applicable to DFTS-OFDM transmission (see Figure 4.11).

4.3.3 User multiplexing with DFTS-OFDM

As mentioned above, by dynamically adjusting the transmitter DFT size and, consequently, also the size of the block of modulation symbols $a_0, a_1, \ldots, a_{M-1}$, the nominal bandwidth of the DFTS-OFDM signal can also be dynamically adjusted. Furthermore, by shifting the IDFT inputs to which the DFT outputs are mapped, the exact frequency-domain "position" of the signal to be transmitted can be adjusted. By these means, DFTS-OFDM allows for uplink FDMA with flexible bandwidth assignment, as illustrated in Figure 4.12.

Copyright Elsevier 2024

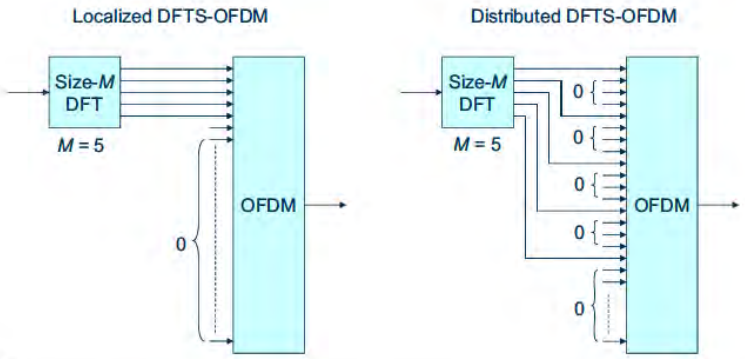


FIGURE 4.13

Localized DFTS-OFDM vs. distributed DFTS-OFDM



FIGURE 4.14

Spectrum of localized and distributed DFTS-OFDM signals

Figure 4.12a illustrates the case of multiplexing the transmissions from two terminals with equal bandwidth assignments—that is, equal DFT sizes-M—while Figure 4.12b illustrates the case of differently sized bandwidth assignments.

4.3.4 Distributed DFTS-OFDM

What has been illustrated in Figure 4.8 can more specifically be referred to as *Localized DFTS-OFDM*, referring to the fact that the output of the DFT is mapped to *consecutive* inputs of the OFDM modulator. An alternative is to map the output of the DFT to *equidistant* inputs of the OFDM modulator with zeros inserted in between, as illustrated in Figure 4.13. This can also be referred to as *Distributed DFTS-OFDM*.

Figure 4.14 illustrates the basic structure of the transmitted spectrum for localized and distributed DFTS-OFDM respectively. Although the spectrum of the localized DFTS-OFDM signal clearly indicates a single-carrier transmission, this is not as clearly seen from the spectrum of the distributed DFTS-OFDM signal. However, it can be shown that a distributed DFTS-OFDM signal has similar power variations as a localized DFTS-OFDM. Actually, it can be shown that a distributed DFTS-OFDM signal is equivalent to so-called *Interleaved FDMA* (IFDMA) [28]. The benefit of distributed DFTS-OFDM, compared to localized

Copyright Elsevier 2024



FIGURE 4.15

User multiplexing in the case of localized and distributed DFTS-OFDM

DFTS-OFDM, is the possibility of additional frequency diversity as even a low-rate distributed DFTS-OFDM signal (small DFT size-M) can be spread over a potentially very large overall transmission bandwidth.

User multiplexing in the frequency domain as well as flexible bandwidth allocation is also possible in the case of distributed DFTS-OFDM. However, the different users are interleaved in the frequency domain, as illustrated in the right part of Figure 4.15 (thus the alternative term "*Interleaved FDMA*"). As a consequence, distributed DFTS-OFDM is more sensitive to frequency errors and has higher requirements on power control, compared to localized DFTS-OFDM. This is similar to the case of localized OFDM vs. distributed OFDM as discussed in Section 3.10.

Copyright Elsevier 2024

CHAPTER

Multi-Antenna Techniques

5

Multi-antenna techniques can be seen as a joint name for a set of techniques with the common theme that they rely on the use of multiple antennas at the receiver and/or the transmitter, in combination with more or less advanced signal processing. Multi-antenna techniques can be used to achieve improved system performance, including improved system capacity (more users per cell) and improved coverage (possibility for larger cells), as well as improved service provisioning—for example, higher per-user data rates. This chapter will provide a general overview of different multi-antenna techniques. How multi-antenna techniques are specifically applied to LTE is discussed in more detail in Chapters 10 and 11.

5.1 Multi-antenna configurations

An important characteristic of any multi-antenna configuration is the distance between the different antenna elements, to a large extent due to the relation between the antenna distance and the mutual correlation between the radio-channel fading experienced by the signals at the different antennas.

The antennas in a multi-antenna configuration can be located relatively far from each other, typically implying a relatively low mutual correlation. Alternatively, the antennas can be located relatively close to each other, typically implying a high mutual fading correlation—in essence, the different antennas experience the same, or at least very similar, instantaneous fading. Whether high or low correlation is desirable depends on what is to be achieved with the multi-antenna configuration (diversity, beam-forming, or spatial multiplexing), as discussed further below.

What actual antenna distance is needed for low (alternatively high) fading correlation depends on the wavelength or, equivalently, the carrier frequency used for the radio communication. However, it also depends on the deployment scenario.

In the case of base-station antennas in typical macro-cell environments (relatively large cells, relatively high base-station antenna positions, etc.), an antenna distance of the order of 10 wavelengths is typically needed to ensure a low mutual fading correlation. At the same time, for a terminal in the same kind of environment, an antenna distance of the order of only half a wavelength (0.5λ) is often sufficient to achieve relatively low mutual correlation [29]. The reason for the difference between the base station and the terminal in this respect is that, in the macro-cell scenario, the multi-path reflections that cause the fading mainly occur in the

4G: LTE/LTE-Advanced for Mobile Broadband. http://dx.doi.org/10.1016/B978-0-12-419985-9.00005-2
Copyright © 2014 Erik Dahlman, Stefan Parkvall and Johan Sköld. Published by Elsevier Ltd. All rights reserved.

Copyright Elsevier 2024

Samsung Exhibit 1017, Page 89 of 538
Samsung Electronics Co., Ltd. v. Asus Technology Licensing Inc.
IPR2024-01217

near-zone around the terminal. Thus, as seen from the terminal, the different paths will typically arrive from a wide angle, implying a low fading correlation already with a relatively small antenna distance. At the same time, as seen from the (macro-cell) base station the different paths will typically arrive within a much smaller angle, implying the need for significantly larger antenna distance to achieve low fading correlation.[1]

On the other hand, in other deployment scenarios, such as micro-cell deployments with base-station antennas below rooftop level and indoor deployments, the environment as seen from the base station is more similar to the environment as seen from the terminal. In such scenarios, a smaller base-station antenna distance is typically sufficient to ensure relatively low mutual correlation between the fading experienced by the different antennas.

The above discussion assumed antennas with the same polarization direction. Another means to achieve low mutual fading correlation is to apply different polarization directions for the different antennas [29]. The antennas can then be located relatively close to each other, implying a compact antenna arrangement, while still experiencing low mutual fading correlation.

5.2 Benefits of multi-antenna techniques

The availability of multiple antennas at the transmitter and/or the receiver can be utilized in different ways to achieve different aims:

- Multiple antennas at the transmitter and/or the receiver can be used to provide additional diversity against fading on the radio channel. In this case, the channels experienced by the different antennas should have low mutual correlation, implying the need for a sufficiently large inter-antenna distance (*spatial diversity*), or the use of different antenna polarization directions (*polarization diversity*).
- Multiple antennas at the transmitter and/or the receiver can be used to "shape" the overall antenna beam (transmit beam and receive beam respectively) in a certain way—for example, to maximize the overall antenna gain in the direction of the target receiver/transmitter or to suppress specific dominant interfering signals. Such *beam-forming* can be based either on high or low fading correlation between the antennas, as is further discussed in Section 5.4.2.
- The simultaneous availability of multiple antennas at the transmitter and the receiver can be used to create what can be seen as multiple parallel communication "channels" over the radio interface. This provides the possibility for very high bandwidth utilization without a corresponding reduction in power efficiency or, in other words, the possibility for very high data rates within a limited bandwidth without a disproportionately large degradation in terms of coverage. Herein we will refer to this as *spatial multiplexing.* It is

[1]Although the term "arrive" is used above, the situation is exactly the same in the case of multiple *transmit* antennas. Thus, in the case of multiple base-station transmit antennas in a macro-cell scenario, the antenna distance typically needs to be a number of wavelengths to ensure low fading correlation while, in the case of multiple transmit antennas at the terminal, an antenna distance of a fraction of a wavelength is sufficient.

Copyright Elsevier 2024

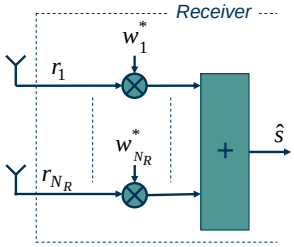


FIGURE 5.1

Linear receive-antenna combining

often also referred to as MIMO (Multi-Input Multi-Output) antenna processing. Spatial multiplexing is discussed in Section 5.5.

5.3 Multiple receive antennas

Perhaps the most straightforward and historically the most commonly used multi-antenna configuration is multiple antennas at the receiver side. This is often referred to as *receive diversity* or *Rx diversity* even if the aim of the multiple receive antennas is not always to achieve additional diversity against radio-channel fading.

Figure 5.1 illustrates the basic principle of linear combination of signals r_1, …, r_{N_R} received at N_R different antennas, with the received signals being multiplied by complex weight factors $w_1^* \ldots w_{N_R}^*$ before being added together. In vector notation this *linear receive-antenna combination* can be expressed as:[2]

$$\hat{s} = \left[w_1^* \ldots w_{N_R}^* \right] \cdot \begin{bmatrix} r_1 \\ \vdots \\ r_{N_R} \end{bmatrix} = \overline{w}^H \cdot \overline{r} \tag{5.1}$$

What is outlined in (5.1) and Figure 5.1 is linear receive-antenna combination in general. Different specific antenna-combining approaches vary in the exact choice of the weight vector $\overline{w}$.

Assuming that the transmitted signal is only subject to non-frequency-selective fading and white noise—that is, there is no radio-channel time dispersion—the signals received at the different antennas in Figure 5.1 can be expressed as:

$$\overline{r} = \begin{pmatrix} r_1 \\ \vdots \\ r_{N_R} \end{pmatrix} = \begin{pmatrix} h_1 \\ \vdots \\ h_{N_R} \end{pmatrix} \cdot s + \begin{pmatrix} n_1 \\ \vdots \\ n_{N_R} \end{pmatrix} = \overline{h} \cdot s + \overline{n} \tag{5.2}$$

[2]Note that the weight factors are expressed as complex conjugates of w_1, …, w_{N_R}.

Copyright Elsevier 2024

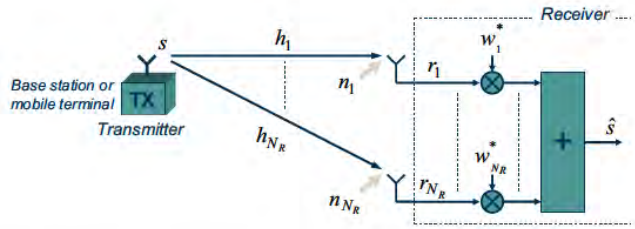


FIGURE 5.2

Linear receive-antenna combining

where s is the transmitted signal, the vector $\overline{h}$ consists of the N_R complex channel gains, and the vector $\overline{n}$ consists of the noise impairing the signals received at the different antennas (see also Figure 5.2).

One can easily show that, to maximize the signal-to-noise ratio after linear combination, the weight vector $\overline{w}$ should be selected as [12]:

$$\overline{w}_{MRC} = \overline{h} \tag{5.3}$$

This is also known as *Maximum-Ratio Combination* (MRC). The MRC weights fulfill two purposes:

- phase rotate the signals received at the different antennas to compensate for the corresponding channel phases and ensure that the signals are phase aligned when added together (coherent combination);
- weight the signals in proportion to their corresponding channel gains—that is, apply higher weights for stronger received signals.

In the case of mutually uncorrelated antennas—sufficiently large antenna distances or different polarization directions—the channel gains $h_1, \ldots, h_{N_R}$ are uncorrelated and the linear antenna combination provides diversity of order N_R. In terms of receiver-side beam-forming, selecting the antenna weights according to (5.3) corresponds to a receiver beam with maximum gain N_R in the direction of the target signal. Thus, the use of multiple receive antennas may increase the post-combiner signal-to-noise ratio in proportion to the number of receive antennas.

MRC is an appropriate antenna-combining strategy when the received signal is mainly impaired by noise. However, in many cases of mobile communication the received signal is mainly impaired by interference from other transmitters within the system, rather than by noise. In a situation with a relatively large number of interfering signals of approximately equal strength, maximum-ratio combination is typically still a good choice as, in this case, the overall interference will appear relatively "noise-like" with no specific direction of arrival. However, in situations where there is a single dominating interferer (or, in the general case, a limited number of dominating interferers), as illustrated in Figure 5.3, improved performance can be achieved if, instead of selecting the antenna weights to maximize the received signal-to-noise ratio after

Copyright Elsevier 2024

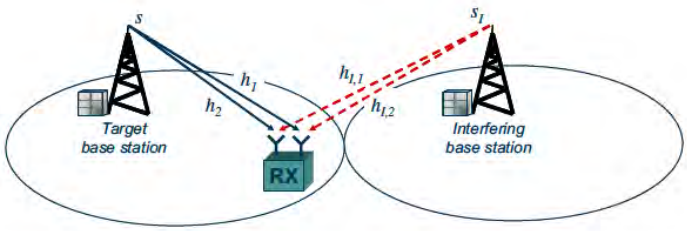


FIGURE 5.3

Downlink scenario with a single dominating interferer (special case of only two receive antennas)

antenna combination (MRC), the antenna weights are selected so that the interferer is suppressed. In terms of receiver-side beam-forming this corresponds to a receiver beam with high attenuation in the direction of the interferer, rather than focusing the receiver beam in the direction of the target signal. Applying receive-antenna combination with a target to suppress specific interferers is often referred to as *Interference Rejection Combination* (IRC) [30].

In the case of a single dominating interferer as outlined in Figure 5.3, expression (5.2) can be extended according to:

$$\overline{r} = \begin{pmatrix} r_1 \\ \vdots \\ r_{N_R} \end{pmatrix} = \begin{pmatrix} h_1 \\ \vdots \\ h_{N_R} \end{pmatrix} \cdot s + \begin{pmatrix} h_{I,1} \\ \vdots \\ h_{I,N_R} \end{pmatrix} \cdot s_I + \begin{pmatrix} n_1 \\ \vdots \\ n_{N_R} \end{pmatrix} = \overline{h} \cdot s + \overline{h}_I \cdot s_I + \overline{n} \qquad (5.4)$$

where s_I is the transmitted interferer signal and the vector $\overline{h}_I$ consists of the complex channel gains from the interferer to the N_R receive antennas. By applying expressions (5.1)–(5.4) it is clear that the interfering signal will be completely suppressed if the weight vector $\overline{w}$ is selected to fulfill the expression:

$$\overline{w}^H \cdot \overline{h}_I = 0 \qquad (5.5)$$

In the general case, (5.5) has $N_R - 1$ non-trivial solutions, indicating flexibility in the weight-vector selection. This flexibility can be used to suppress additional dominating interferers. More specifically, in the general case of N_R receive antennas there is a possibility to, at least in theory, completely suppress up to $N_R - 1$ separate interferers. However, such a choice of antenna weights, providing complete suppression of a number of dominating interferers, may lead to a large, potentially very large, increase in the noise level after the antenna combination. This is similar to the potentially large increase in the noise level in the case of a Zero-Forcing equalizer as discussed in Chapter 4.

Thus, similar to the case of linear equalization, a better approach is to select the antenna weight vector $\overline{w}$ to minimize the mean square error:

$$\varepsilon = E\left\{ \left| \hat{s} - s \right|^2 \right\} \qquad (5.6)$$

also known as *Minimum Mean Square Error* (MMSE) combination [22,29].

Copyright Elsevier 2024

Samsung Exhibit 1017, Page 93 of 538
Samsung Electronics Co., Ltd. v. Asus Technology Licensing Inc.
IPR2024-01217

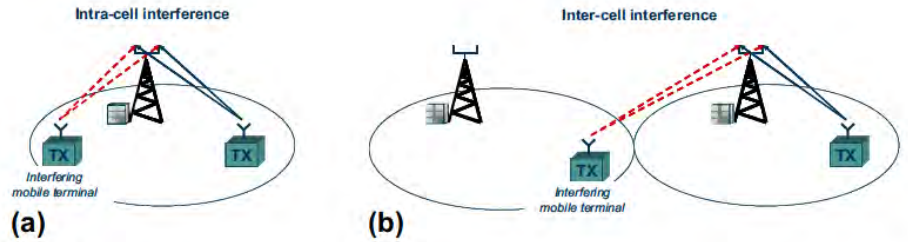


FIGURE 5.4

Receiver scenario with one strong interfering terminal: (a) intra-cell interference and (b) inter-cell interference

Although Figure 5.3 illustrates a downlink scenario with a dominating interfering base station, IRC can also be applied to the uplink to suppress interference from specific terminals. In this case, the interfering terminal may either be in the same cell as the target terminal (*intra-cell interference*) or in a neighboring cell (*inter-cell interference*) (see Figure 5.4a and b respectively). Suppression of intra-cell interference is relevant in the case of a non-orthogonal uplink—that is, when multiple terminals are transmitting simultaneously using the same time–frequency resource. Uplink intra-cell-interference suppression by means of IRC is sometimes also referred to as *Spatial-Division Multiple Access* (SDMA) [31,32].

As discussed in Chapter 2, in practice a radio channel is always subject to at least some degree of time dispersion or, equivalently, frequency selectivity, causing corruption to a wideband signal. As discussed in Chapter 4, one method to counteract such signal corruption is to apply linear equalization, either in the time or frequency domain.

It should be clear from the discussion above that linear receive-antenna combination has many similarities to linear equalization:

- Linear time-domain (frequency-domain) filtering/equalization as described in Chapter 4 implies that linear processing is applied to signals received at different time instances (different frequencies) with a target to maximize the post-equalizer SNR (MRC-based equalization), or to suppress signal corruption due to radio-channel frequency selectivity (zero-forcing equalization, MMSE equalization, etc.).
- Linear receive-antenna combination implies that linear processing is applied to signals received at different antennas—that is, processing in the spatial domain—with a target to maximize the post-combiner SNR (MRC-based combination), or to suppress specific interferers (IRC based on e.g. MMSE).

Thus, in the general case of frequency-selective channel and multiple receive antennas, two-dimensional time/space linear processing/filtering can be applied as illustrated in Figure 5.5, where the linear filtering can be seen as a generalization of the antenna weights of

Copyright Elsevier 2024

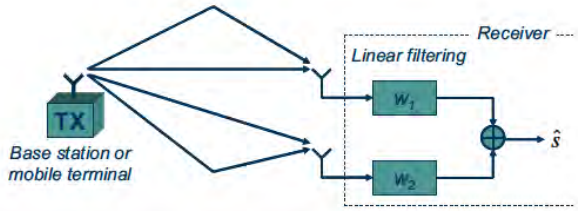


FIGURE 5.5

Two-dimensional space/time linear processing (two receive antennas)

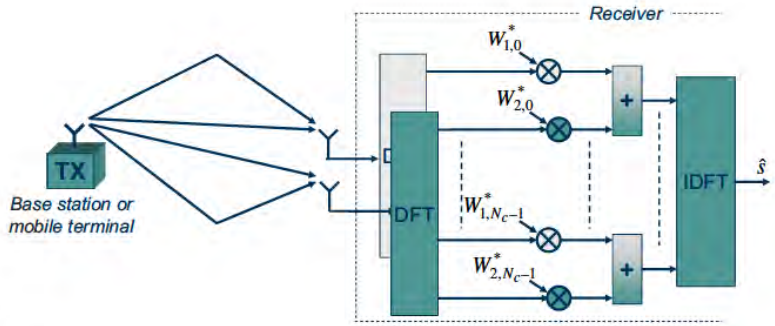


FIGURE 5.6

Two-dimensional space/frequency linear processing (two receive antennas)

Figure 5.1. The filters should be jointly selected to minimize the overall impact of noise, interference, and signal corruption due to radio-channel frequency selectivity.

Alternatively, especially in the case when cyclic-prefix insertion has been applied at the transmitter side, two-dimensional frequency/space linear processing can be applied, as illustrated in Figure 5.6. The frequency-domain weights should then be jointly selected to minimize the overall impact of noise, interference, and signal corruption due to radio-channel frequency selectivity.

The frequency/space processing outlined in Figure 5.6, without the IDFT, is also applicable if receive diversity is to be applied to OFDM transmission. In the case of OFDM transmission, there is no signal corruption due to radio-channel frequency selectivity. Thus, the frequency-domain coefficients of Figure 5.6 can be selected taking into account only noise and interference. In principle, this means the antenna-combining schemes discussed above (MRC and IRC) are applied on a *per-subcarrier* basis.

Note that, although Figures 5.5 and 5.6 assume two receive antennas, the corresponding receiver structures can straightforwardly be extended to more than two antennas.

Copyright Elsevier 2024

5.4 Multiple transmit antennas

As an alternative, or complement, to multiple receive antennas, diversity and beam-forming can also be achieved by applying multiple antennas at the transmitter side. The use of multiple transmit antennas is primarily of interest for the downlink—that is, at the base station. In this case, the use of multiple transmit antennas provides an opportunity for diversity and beam-forming without the need for additional receive antennas and corresponding additional receiver chains at the terminal. On the other hand, due to complexity the use of multiple transmit antennas for the uplink—that is, at the terminal—is less attractive. In this case, it is typically preferred to apply additional receive antennas and corresponding receiver chains at the base station.

5.4.1 Transmit-antenna diversity

If no knowledge of the downlink channels of the different transmit antennas is available at the transmitter, multiple transmit antennas cannot provide beam-forming, but only diversity. For this to be possible, there should be low mutual correlation between the channels of the different antennas. As discussed in Section 5.1, this can be achieved by means of a sufficiently large distance between the antennas, or by the use of different antenna polarization directions. Assuming such antenna configurations, different approaches can be taken to realize the diversity offered by the multiple transmit antennas.

5.4.1.1 Delay diversity

As discussed in Chapter 2, a radio channel subject to time dispersion, with the transmitted signal propagating to the receiver via multiple, independently fading paths with different delays, provides the possibility for multi-path diversity or, equivalently, frequency diversity. Thus, multi-path propagation is actually beneficial in terms of radio-link performance, assuming that the amount of multi-path propagation is not too extensive and that the transmission scheme includes tools to counteract signal corruption due to the radio-channel frequency selectivity—for example, by means of OFDM transmission or the use of advanced receiver-side equalization.

If the channel in itself is not time dispersive, the availability of multiple transmit antennas can be used to create *artificial time dispersion* or, equivalently, *artificial frequency selectivity* by transmitting identical signals with different relative delays from the different antennas. In this way, the antenna diversity—that is, the fact that the fading experienced by the different antennas have low mutual correlation—can be transformed into frequency diversity. This kind of *delay diversity* is illustrated in Figure 5.7 for the special case of two transmit antennas. The relative delay T should be selected to ensure a suitable amount of frequency selectivity over the bandwidth of the signal to be transmitted. It should be noted that, although Figure 5.7 assumes two transmit antennas, delay diversity can straightforwardly be extended to more than two transmit antennas with different relative delays for each antenna.

Copyright Elsevier 2024

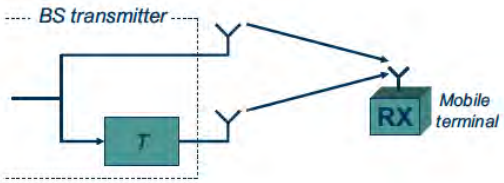


FIGURE 5.7

Two-antenna delay diversity

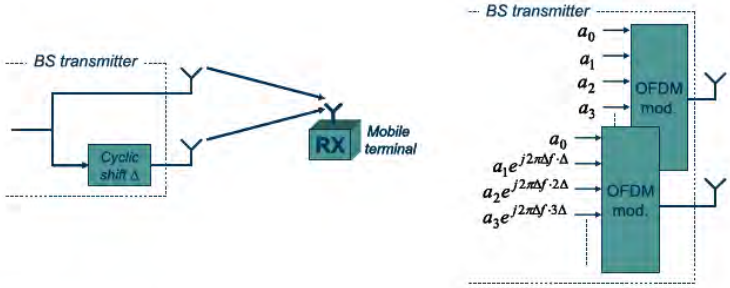


FIGURE 5.8

Two-antenna Cyclic-Delay Diversity (CDD)

Delay diversity is in essence invisible to the terminal, which will simply see a single radio channel subject to additional time dispersion. Delay diversity can straightforwardly be introduced in an existing mobile-communication system without requiring any specific support in a corresponding radio-interface standard. Delay diversity is also applicable to basically any kind of transmission scheme that is designed to handle and benefit from frequency-selective fading.

5.4.1.2 *Cyclic-delay diversity*

Cyclic-Delay Diversity (CDD) [33] is similar to delay diversity with the main difference that cyclic-delay diversity operates block-wise and applies *cyclic shifts*, rather than linear delays, to the different antennas (see Figure 5.8). Thus, cyclic-delay diversity is applicable to block-based transmission schemes such as OFDM and DFTS-OFDM.

In the case of OFDM transmission, a cyclic shift of the time-domain signal corresponds to a frequency-dependent phase shift before OFDM modulation, as illustrated in Figure 5.8b. Similar to delay diversity, this will create artificial frequency selectivity as seen by the receiver.

Also similar to delay diversity, CDD can straightforwardly be extended to more than two transmit antennas with different cyclic shifts for each antenna.

Copyright Elsevier 2024

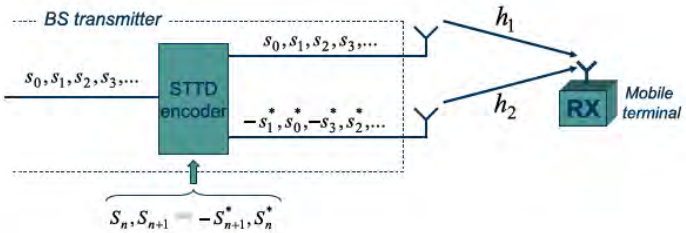


FIGURE 5.9

Space Time Transmit Diversity (STTD)

5.4.2 Means of space–time coding

Space–time coding is a general term used to indicate multi-antenna transmission schemes where modulation symbols are mapped in the time and spatial (transmit-antenna) domain to capture the diversity offered by the multiple transmit antennas. Two-antenna space–time block coding (STBC), more specifically a scheme referred to as *Space–Time Transmit Diversity* (STTD), has been part of the 3G WCDMA standard already from its first release [34].

As shown in Figure 5.9, STTD operates on pairs of modulation symbols. The modulation symbols are directly transmitted on the first antenna. However, on the second antenna the order of the modulation symbols within a pair is reversed. Furthermore, the modulation symbols are sign-reversed and complex-conjugated, as illustrated in Figure 5.9.

In vector notation, STTD transmission can be expressed as:

$$\bar{r} = \begin{pmatrix} r_{2n} \\ r_{2n+1}^* \end{pmatrix} = \begin{pmatrix} h_1 & -h_2 \\ h_2^* & h_1^* \end{pmatrix} \cdot \begin{pmatrix} s_{2n} \\ s_{2n+1}^* \end{pmatrix} = H \cdot \bar{s} \tag{5.7}$$

where r_{2n} and r_{2n+1} are the received symbols in absence of noise during the symbol intervals $2n$ and $2n+1$ respectively.[3] It should be noted that this expression assumes that the channel coefficients h_1 and h_2 are constant over the time corresponding to two consecutive symbol intervals, an assumption that is typically valid. As the matrix H is a scaled unitary matrix, the sent symbols s_{2n} and s_{2n+1} can be recovered from the received symbols r_{2n} and r_{2n+1}, without any interference between the symbols, by applying the matrix $W = H^{-1}$ to the vector $\bar{r}$.

The two-antenna space–time coding of Figure 5.9 can be said to be of rate one, implying that the input symbol rate is the same as the symbol rate at each antenna, corresponding to a bandwidth utilization of 1. Space–time coding can also be extended to more than two antennas. However, in the case of complex-valued modulation, such as QPSK or 16/64QAM, space–time codes of rate one without any inter-symbol interference (*orthogonal space–time codes*) only exist for two antennas [35]. If inter-symbol interference is to be avoided in the case of more than two antennas, space–time codes with rate less than one must be used, corresponding to reduced bandwidth utilization.

[3]Note that, for convenience, complex conjugates have been applied for the second elements of $\bar{r}$ and $\bar{s}$.

Copyright Elsevier 2024

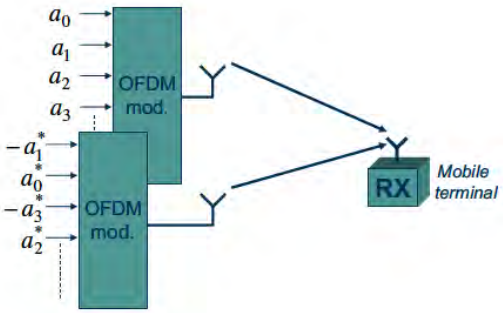


FIGURE 5.10

Space Frequency Transmit Diversity assuming two transmit antennas

5.4.2.1 *Diversity by means of space–frequency coding*

Space–frequency block coding (SFBC) is similar to space–time block coding, with the difference that the encoding is carried out in the antenna/frequency domains rather than in the antenna/time domains. Thus, space–frequency coding is applicable to OFDM and other "frequency-domain" transmission schemes. The space–frequency equivalence to STTD (which could also be referred to as Space–Frequency Transmit Diversity, SFTD) is illustrated in Figure 5.10. The block of (frequency-domain) modulation symbols a_0, a_1, a_2, a_3, … is directly mapped to OFDM carriers of the first antenna, while the block of symbols $-a_1^*, a_0^*, -a_3^*, a_2^*$ is mapped to the corresponding subcarriers of the second antenna.

Similar to space–time coding, the drawback of space–frequency coding is that there is no straightforward extension to more than two antennas unless a rate reduction is acceptable.

Comparing Figure 5.10 with the right-hand side of Figure 5.8, it can be noted that the difference between SFBC and two-antenna cyclic-delay diversity in essence lies in how the block of frequency-domain modulation symbols are mapped to the second antenna. The benefit of SFBC compared to CDD is that SFBC provides diversity at modulation-symbol level while CDD, in the case of OFDM, must rely on channel coding in combination with frequency-domain interleaving to provide diversity.

5.4.3 Transmitter-side beam-forming

If some knowledge of the downlink channels of the different transmit antennas (more specifically, some knowledge of the relative channel phases) is available at the transmitter side, multiple transmit antennas can, in addition to diversity, also provide beam-forming—the shaping of the overall antenna beam in the direction of a target receiver. In general, such beam-forming can increase the signal strength at the receiver with up to a factor N_T—that is, in proportion to the number of transmit antennas. When discussing transmission schemes

Copyright Elsevier 2024

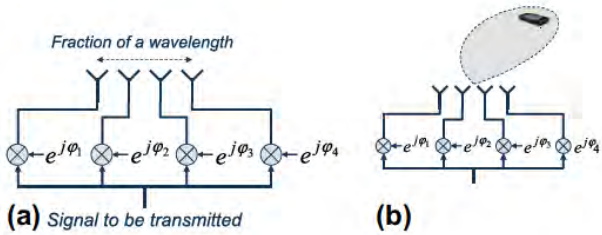


FIGURE 5.11

Classical beam-forming with high mutual antennas correlation: (a) antenna configuration and (b) beam-structure

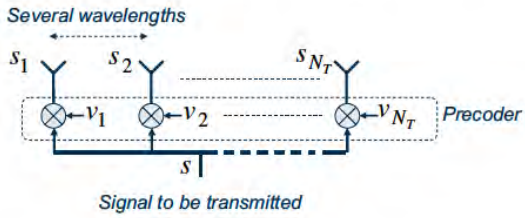


FIGURE 5.12

Precoder-based beam-forming in case of low mutual antenna correlation

relying on multiple transmit antennas to provide beam-forming, one can distinguish between the cases of *high* and *low mutual antenna correlation* respectively.

High mutual antenna correlation typically implies an antenna configuration with a small inter-antenna distance, as illustrated in Figure 5.11a. In this case, the channels between the different antennas and a specific receiver are essentially the same, including the same radio-channel fading, except for a direction-dependent phase difference. The overall transmission beam can then be steered in different directions by applying *different phase shifts* to the signals to be transmitted on the different antennas, as illustrated in Figure 5.11b.

This approach to transmitter-side beam-forming, with different phase shifts applied to highly correlated antennas, is sometimes referred to as "classical" beam-forming. Due to the small antenna distance, the overall transmission beam will be relatively wide and any adjustments of the beam direction—in practice, adjustments of the antenna phase shifts—will typically be carried out on a relatively slow basis. The adjustments could, for example, be based on estimates of the direction to the target terminal derived from uplink measurements. Furthermore, due to the assumption of high correlation between the different transmit antennas, classical beam-forming cannot provide any diversity against radio-channel fading but only an increase of the received signal strength.

Low mutual antenna correlation typically implies either a sufficiently large antenna distance, as illustrated in Figure 5.12, or different antenna polarization directions. With low

Copyright Elsevier 2024

mutual antenna correlation, the basic beam-forming principle is similar to that of Figure 5.11—that is, the signals to be transmitted on the different antennas are multiplied by different complex weights. However, in contrast to classical beam-forming, the antenna weights should now take general complex values and both the phase and the amplitude of the signals to be transmitted on the different antennas can be adjusted. This reflects the fact that, due to the low mutual antenna correlation, both the phase and the instantaneous gain of the channels of each antenna may differ.

Applying different complex weights to the signals to be transmitted on the different antennas can be expressed, in vector notation, as applying a *precoding vector* $\overline{v}$ to the signal to be transmitted, according to:

$$\overline{s} = \begin{pmatrix} s_1 \\ \vdots \\ s_{N_T} \end{pmatrix} = \begin{pmatrix} v_1 \\ \vdots \\ v_{N_T} \end{pmatrix} \cdot s = \overline{v} \cdot s \tag{5.8}$$

It should be noted that classical beam-forming according to Figure 5.11 can also be described according to (5.8)—that is, as *transmit-antenna precoding*—with the constraint that the antenna weights are limited to unit gain and only provide phase shifts to the different transmit antennas.

Assuming that the signals transmitted from the different antennas are only subject to non-frequency-selective fading and white noise—that is, there is no radio-channel time dispersion—it can be shown [36] that, in order to maximize the received signal power, the precoding weights should be selected according to:

$$v_i = \frac{h_i^*}{\sqrt{\sum_{k=1}^{N_T} |h_k|^2}} \tag{5.9}$$

That is, as the complex conjugate of the corresponding channel coefficient h_i and with a normalization to ensure a fixed overall transmit power. The precoding vector thus:

- phase rotates the transmitted signals to compensate for the instantaneous channel phase and ensure that the received signals are received phase aligned;
- allocates power to the different antennas with, in general, more power being allocated to antennas with good instantaneous channel conditions (high channel gain $|h_i|$);
- ensures an overall unit (or any other constant) transmit power.

A key difference between classical beam-forming according to Figure 5.11, assuming high mutual antenna correlation, and beam-forming according to Figure 5.12, assuming low mutual antenna correlation, is that, in the latter case, there is a need for more detailed channel knowledge, including estimates of the instantaneous channel fading. Updates to the precoding vector are thus typically done on a relatively short time scale to capture the fading variations. As the adjustment of the precoder weights takes into account the instantaneous fading, including the instantaneous channel gain, fast beam-forming according to Figure 5.12 also provides diversity against radio-channel fading.

Copyright Elsevier 2024

Samsung Exhibit 1017, Page 101 of 538
Samsung Electronics Co., Ltd. v. Asus Technology Licensing Inc.
IPR2024-01217

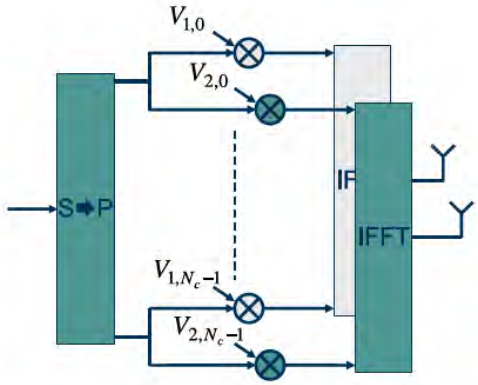


FIGURE 5.13

Per-subcarrier precoding in case of OFDM (two transmit antennas)

Furthermore, at least in the case of communication based on *Frequency-Division Duplex* (FDD), with uplink and downlink communication taking place in different frequency bands, the fading is typically uncorrelated between the downlink and uplink. Thus, in the case of FDD, only the terminal can determine the downlink fading. The terminal would then report an estimate of the downlink channel to the base station by means of uplink signaling. Alternatively, the terminal may, in itself, select a suitable precoding vector from a limited set of possible precoding vectors, the so-called precoder codebook, and report this to the base station.

On the other hand, in the case of *Time-Division Duplex* (TDD), with uplink and downlink communication taking place in the same frequency band but in separate non-overlapping time slots, there is typically a high fading correlation between the downlink and uplink. In this case, the base station could, at least in theory, determine the instantaneous downlink fading from measurements on the uplink, thus avoiding the need for any feedback. Note, however, that this assumes that the terminal is continuously transmitting on the uplink.

The above discussion assumed that the channel gain was constant in the frequency domain—that is, there was no radio-channel frequency selectivity. In the case of a frequency-selective channel there is not a single channel coefficient per antenna, based on which the antenna weights can be selected according to (5.9). However, in the case of OFDM transmission, each subcarrier will typically experience a frequency-non-selective channel, thus the precoding of Figure 5.12 can be carried out on a per-subcarrier basis as outlined in Figure 5.13, where the precoding weights of each subcarrier should be selected according to (5.9).

It should be pointed out that in the case of single-carrier transmission the one-weight-per-antenna approach outlined in Figure 5.12 can be extended to also take into account a time-dispersive/frequency-selective channel [37].

Copyright Elsevier 2024

5.5 Spatial multiplexing

The use of multiple antennas at both the transmitter and the receiver can simply be seen as a tool to further improve the signal-to-noise/interference ratio and/or achieve additional diversity against fading, compared to the use of only multiple receive antennas or multiple transmit antennas. However, in the case of multiple antennas at both the transmitter and the receiver there is also the possibility for so-called *spatial multiplexing*, allowing for more efficient utilization of high signal-to-noise/interference ratios and significantly higher data rates over the radio interface.

5.5.1 Basic principles

It should be clear from the previous sections that multiple antennas at the receiver and the transmitter can be used to improve the receiver signal-to-noise ratio in proportion to the number of antennas by applying beam-forming at the receiver and the transmitter. In the general case of N_T transmit antennas and N_R receive antennas, the receiver signal-to-noise ratio can be made to increase in proportion to the product $N_T \times N_R$. As discussed in Chapter 2, such an increase in the receiver signal-to-noise ratio allows for a corresponding increase in the achievable data rates, assuming that the data rates are power limited rather than bandwidth limited. However, once the bandwidth-limited range of operation is reached, the achievable data rates start to saturate unless the bandwidth is also allowed to increase.

One way to understand this saturation in achievable data rates is to consider the basic expression for the normalized channel capacity:

$$\frac{C}{BW} = \log_2\left(1 + \frac{S}{N}\right) \tag{5.10}$$

where, by means of beam-forming, the signal-to-noise ratio S/N can be made to grow proportionally to $N_T \times N_R$. In general, $\log_2(1 + x)$ is proportional to x for small x, implying that, for low signal-to-noise ratios, the capacity grows approximately proportionally to the signal-to-noise ratio. However, for larger x, $\log_2(1 + x) \approx \log_2(x)$, implying that, for larger signal-to-noise ratios, capacity grows only logarithmically with the signal-to-noise ratio.

In the case of multiple antennas at the transmitter *and* the receiver it is, *under certain conditions*, possible to create up to $N_L = \min\{N_T, N_R\}$ parallel "channels" each with N_L times lower signal-to-noise ratio (the signal power is "split" between the channels)—that is, with a channel capacity:

$$\frac{C}{BW} = \log_2\left(1 + \frac{N_R}{N_L} \cdot \frac{S}{N}\right) \tag{5.11}$$

Copyright Elsevier 2024

Samsung Exhibit 1017, Page 103 of 538
Samsung Electronics Co., Ltd. v. Asus Technology Licensing Inc.
IPR2024-01217

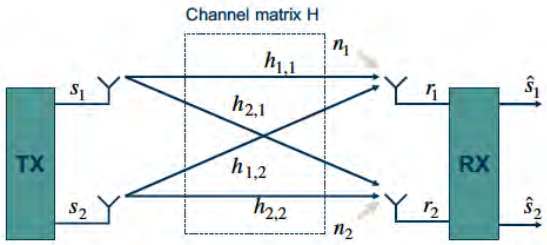


FIGURE 5.14

2 × 2 antenna configuration

As there are now N_L parallel channels, each with a channel capacity given by (5.11), the overall channel capacity for such a multi-antenna configuration is given by:

$$\begin{aligned}\frac{C}{BW} &= \mathrm{N_L}\cdot\log_2\left(1+\frac{\mathrm{N_R}}{\mathrm{N_L}}\cdot\frac{S}{N}\right)\\ &= \min\left\{\mathrm{N_T},\mathrm{N_R}\right\}\cdot\log_2\left(1+\frac{\mathrm{N_R}}{\min\{\mathrm{N_T},\mathrm{N_R}\}}\cdot\frac{S}{N}\right)\end{aligned} \tag{5.12}$$

Thus, *under certain conditions*, the channel capacity can be made to grow essentially linearly with the number of antennas, avoiding the saturation in the data rates. We will refer to this as *Spatial Multiplexing*. The term MIMO (Multiple-Input/Multiple-Output) antenna processing is also very often used, although the term strictly speaking refers to all cases of multiple transmit antennas and multiple receive antennas, including the case of combined transmit and receive diversity.[4]

To understand the basic principles of how multiple parallel channels can be created in the case of multiple antennas at the transmitter and the receiver, consider a 2 × 2 antenna configuration—that is, two transmit antennas and two receive antennas—as outlined in Figure 5.14. Furthermore, assume that the transmitted signals are only subject to non-frequency-selective fading and white noise—there is no radio-channel time dispersion.

Based on Figure 5.14, the received signals can be expressed as:

$$\overline{r}=\begin{pmatrix}r_1\\r_2\end{pmatrix}=\begin{pmatrix}h_{1,1}&h_{1,2}\\h_{2,1}&h_{2,2}\end{pmatrix}\cdot\begin{pmatrix}s_1\\s_2\end{pmatrix}+\begin{pmatrix}n_1\\n_2\end{pmatrix}=H\cdot\overline{s}+\overline{n} \tag{5.13}$$

where H is the 2 × 2 *channel matrix*. This expression can be seen as a generalization of (5.2) in Section 5.3 to multiple transmit antennas, with different signals being transmitted from the different antennas.

[4]The case of a single transmit antenna and multiple receive antennas is, consequently, often referred to as SIMO (Single-Input/Multiple-Output). Similarly, the case of multiple transmit antennas and a single receiver antenna can be referred to as MISO (Multiple-Input/Single-Output).

Copyright Elsevier 2024

Samsung Exhibit 1017, Page 104 of 538
Samsung Electronics Co., Ltd. v. Asus Technology Licensing Inc.
IPR2024-01217

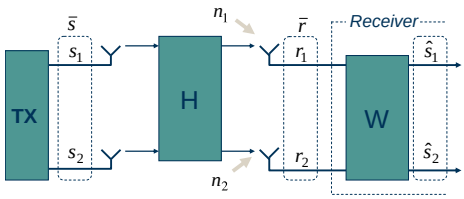


FIGURE 5.15

Linear reception/demodulation of spatially multiplexed signals

Assuming no noise and that the channel matrix H is invertible, the vector $\bar{s}$, and thus both signals s_1 and s_2, can be perfectly recovered at the receiver, with no residual interference between the signals, by multiplying the received vector $\bar{r}$ with a matrix $W = H^{-1}$:

$$\begin{pmatrix} \hat{s}_1 \\ \hat{s}_2 \end{pmatrix} = W \cdot \bar{r} = \begin{pmatrix} s_1 \\ s_2 \end{pmatrix} + H^{-1} \cdot \bar{n} \tag{5.14}$$

Although the vector $\bar{s}$ can be perfectly recovered in the case of no noise, as long as the channel matrix H is invertible, (5.14) also indicates that the properties of H will determine to what extent the joint demodulation of the two signals will increase the noise level. More specifically, the closer the channel matrix is to being a singular matrix, the larger the increase in the noise level.

One way to interpret the matrix W is to realize that the signals transmitted from the two transmit antennas are two signals causing interference to each other. The two receive antennas can then be used to carry out IRC, in essence completely suppressing the interference from the signal transmitted on the second antenna when detecting the signal transmitted at the first antenna and vice versa. The rows of the receiver matrix W simply implement such IRC.

In the general case, a multiple-antenna configuration will consist of N_T transmit antennas and N_R receive antennas. As discussed above, in such a case, the number of parallel signals that can be spatially multiplexed is, at least in practice, upper limited by $N_L = \min\{N_T, N_R\}$. This can intuitively be understood from the fact that:

- No more than N_T different signals can be transmitted from N_T transmit antennas, implying a maximum of N_T spatially multiplexed signals.
- With N_R receive antennas, a maximum of $N_R - 1$ interfering signals can be suppressed, implying a maximum of N_R spatially multiplexed signals.

However, in many cases, the number of spatially multiplexed signals, or the *order of the spatial multiplexing*, will be less than N_L given above:

- In the case of very bad channel conditions (low signal-to-noise ratio) there is no gain of spatial multiplexing as the channel capacity is a linear function of the signal-to-noise

Copyright Elsevier 2024

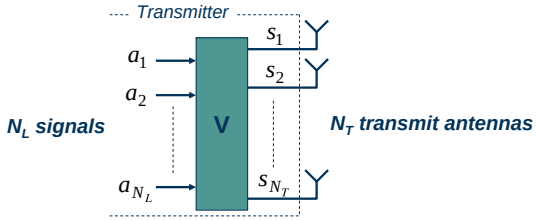


FIGURE 5.16

Precoder-based spatial multiplexing

ratio. In such a case, the multiple transmit and receive antennas should be used for beam-forming to improve the signal-to-noise ratio, rather than for spatial multiplexing.

- In the more general case, the spatial-multiplexing order should be determined based on the properties of the size $N_R \times N_T$ channel matrix. Any excess antennas should then be used to provide beam-forming. Such combined beam-forming and spatial multiplexing can be achieved by means of *precoder-based* spatial multiplexing, as discussed below.

5.5.2 Precoder-based spatial multiplexing

Linear precoding in the case of spatial multiplexing implies that linear processing by means of a size $N_T \times N_L$ precoding matrix is applied at the transmitter side, as illustrated in Figure 5.16. In line with the discussion above, in the general case N_L is equal or smaller than N_T, implying that N_L signals are spatially multiplexed and transmitted using N_T transmit antennas.

It should be noted that precoder-based spatial multiplexing can be seen as a generalization of precoder-based beam-forming as described in Section 5.4.3, with the precoding vector of size $N_T \times 1$ replaced by a precoding matrix of size $N_T \times N_L$.

The precoding of Figure 5.16 can serve two purposes:

- In the case when the number of signals to be spatially multiplexed equals the number of transmit antennas ($N_L = N_T$), the precoding can be used to "orthogonalize" the parallel transmissions, allowing for improved signal isolation at the receiver side.
- In the case when the number of signals to be spatially multiplexed is less than the number of transmit antennas ($N_L < N_T$), the precoding also provides the mapping of the N_L spatially multiplexed signals to the N_T transmit antennas, including the combination of spatially multiplexing and beam-forming.

To confirm that precoding can improve the isolation between the spatially multiplexed signals, express the channel matrix H as its singular-value decomposition [38]:

$$H = W \cdot \Sigma \cdot V^H \tag{5.15}$$

Copyright Elsevier 2024

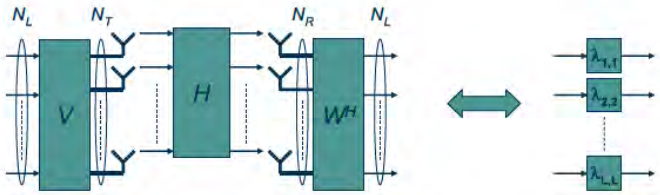


FIGURE 5.17

Orthogonalization of spatially multiplexed signals by means of precoding. $\lambda_{i,i}$ is the ith singular value of the matrix H

where the columns of V and W each form an orthonormal set and Σ is an $N_L \times N_L$ diagonal matrix with the N_L strongest singular values of H as its diagonal elements. By applying the matrix V as precoding matrix at the transmitter side and the matrix WH at the receiver side, one arrives at an equivalent channel matrix $H' = \Sigma$ (see Figure 5.17). As H' is a diagonal matrix, there is no interference between the spatially multiplexed signals at the receiver. At the same time, as both V and W have orthonormal columns, the transmit power as well as the demodulator noise level (assuming spatially white noise) are unchanged.

Clearly, in the case of precoding each received signal will have a certain "quality," depending on the eigenvalues of the channel matrix (see right part of Figure 5.17). This indicates potential benefits of applying dynamic link adaptation in the *spatial domain*—that is, the adaptive selection of the coding rates and/or modulation schemes for each signal to be transmitted.

As the precoding matrix will never perfectly match the channel matrix in practice, there will always be some residual interference between the spatially multiplexed signals. This interference can be taken care of by means of additional receiver-size linear processing according to Figure 5.15 or nonlinear processing as discussed in Section 5.5.3 below.

To determine the precoding matrix V, knowledge about the channel matrix H is needed. Similar to precoder-based beam-forming, a common approach is to have the receiver estimate the channel and decide on a suitable precoding matrix from a set of available precoding matrices (the precoder *codebook*). The receiver then feeds back information about the selected precoding matrix to the transmitter.

5.5.3 Nonlinear receiver processing

The previous sections discussed the use of linear receiver processing to jointly recover spatially multiplexed signals. However, improved demodulation performance can be achieved if nonlinear receiver processing can be applied in the case of spatial multiplexing.

The "optimal" receiver approach for spatially multiplexed signals is to apply *Maximum-Likelihood* (ML) detection [25]. However, in many cases ML detection is too complex to use. Thus, several different proposals have been made for reduced complexity, almost ML, schemes (see, for example, [39]).

Copyright Elsevier 2024

Samsung Exhibit 1017, Page 107 of 538
Samsung Electronics Co., Ltd. v. Asus Technology Licensing Inc.
IPR2024-01217

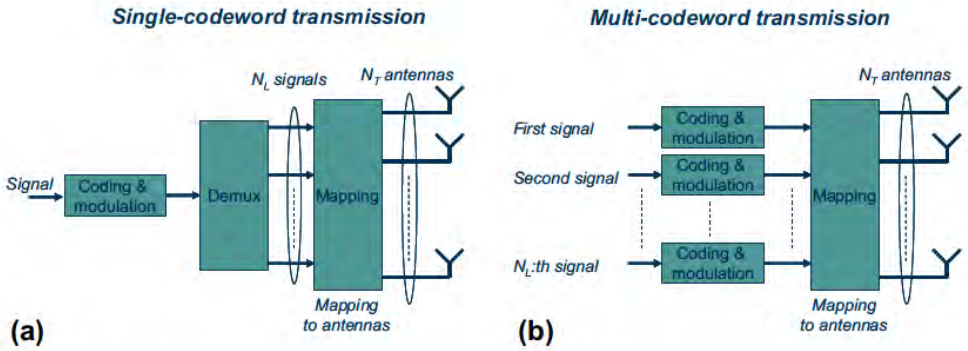


FIGURE 5.18

Single-codeword transmission (a) vs. multi-codeword transmission (b)

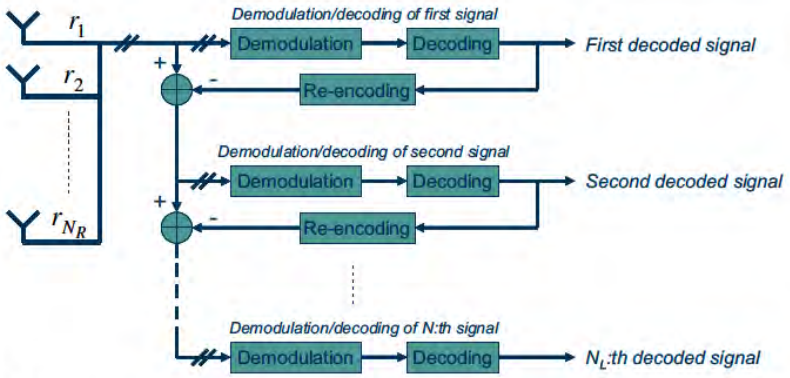


FIGURE 5.19

Demodulation/decoding of spatially multiplexed signals based on successive interference cancellation

Another nonlinear approach to the demodulation of spatially multiplexed signals is to apply *Successive Interference Cancellation* (SIC) [40]. Successive Interference Cancellation is based on an assumption that the spatially multiplexed signals are separately coded before the spatial multiplexing. This is often referred to as *Multi-Codeword* transmission, in contrast to *Single-Codeword* transmission where the spatially multiplexed signals are assumed to be jointly coded (Figure 5.18). It should be understood that, also in the case of multi-codeword transmission, the data may originate from the same source but is then demultiplexed into different signals to be spatially multiplexed before channel coding.

As shown in Figure 5.19, in the case of successive interference cancellation the receiver first demodulates and decodes one of the spatially multiplexed signals. The corresponding

Copyright Elsevier 2024

decoded data is then, if correctly decoded, re-encoded and subtracted from the received signals. A second spatially multiplexed signal can then be demodulated and decoded without, at least in the ideal case, any interference from the first signal—that is, with an improved signal-to-interference ratio. The decoded data of the second signal is then, if correctly decoded, re-encoded and subtracted from the received signal before decoding of a third signal. These iterations continue until all spatially multiplexed signals have been demodulated and decoded.

Clearly, in the case of Successive Interference Cancellation, the first signals to be decoded are subject to higher interference levels, compared to later decoded signals. To work properly, there should be a differentiation in the robustness of the different signals with, at least in principle, the first signal to be decoded being more robust than the second signal, the second signal being more robust than the third signal, etc. Assuming multi-codeword transmission according to Figure 5.18b, this can be achieved by applying different modulation schemes and coding rates to the different signals with, typically, lower-order modulation and lower coding rate, implying a lower data rate, for the first signals to be decoded. This is often referred to as *Per-Antenna Rate Control* (PARC) [41].

Copyright Elsevier 2024

Samsung Exhibit 1017, Page 109 of 538
Samsung Electronics Co., Ltd. v. Asus Technology Licensing Inc.
IPR2024-01217

This page intentionally left blank

Copyright Elsevier 202

CHAPTER

Scheduling, Link Adaptation, and Hybrid ARQ

6

One key characteristic of mobile radio communication is the typically rapid and significant variations in the instantaneous channel conditions. There are several reasons for these variations. Frequency-selective fading will result in rapid and random variations in the channel attenuation. Shadow fading and distance-dependent path loss will also significantly affect the average received signal strength. Finally, the interference at the receiver due to transmissions in other cells and by other terminals will also impact the interference level. Hence, to summarize, there will be rapid, and to some extent random, variations in the experienced quality of each radio link in a cell, variations that must be taken into account and preferably exploited.

In this chapter, some of the techniques for handling variations in the instantaneous radio-link quality will be discussed. *Channel-dependent scheduling* in a mobile-communication system deals with the question of how to share the radio resource(s) available in the system between different users (different terminals) to achieve as efficient resource utilization as possible. Typically, this implies minimizing the amount of resources needed per user and thus allowing for as many users as possible in the system, while still satisfying whatever quality-of-service requirements that may exist. Closely related to scheduling is *link adaptation*, which deals with how to set the transmission parameters of a radio link to handle variations of the radio-link quality.

Both channel-dependent scheduling and link adaptation try to exploit the channel variations through appropriate processing *prior* to transmission of the data. However, due to the random nature of the variations in the radio-link quality, perfect adaptation to the instantaneous radio-link quality is never possible. *Hybrid ARQ*, which requests retransmission of erroneously received data packets, is therefore useful. This can be seen as a mechanism for handling variations in the instantaneous radio-link quality *after* transmission and nicely complements channel-dependent scheduling and link adaptation. Hybrid ARQ also serves the purpose of handling random errors due to, for example, noise in the receiver.

6.1 Link adaptation: power and rate control

Historically, *dynamic transmit-power control* has been used in CDMA-based mobile-communication systems such as WCDMA and CDMA2000 to compensate for variations

4G: LTE/LTE-Advanced for Mobile Broadband. http://dx.doi.org/10.1016/B978-0-12-419985-9.00006-4
Copyright © 2014 Erik Dahlman, Stefan Parkvall and Johan Sköld. Published by Elsevier Ltd. All rights reserved.

Copyright Elsevier 2024

Samsung Exhibit 1017, Page 111 of 538
Samsung Electronics Co., Ltd. v. Asus Technology Licensing Inc.
IPR2024-01217

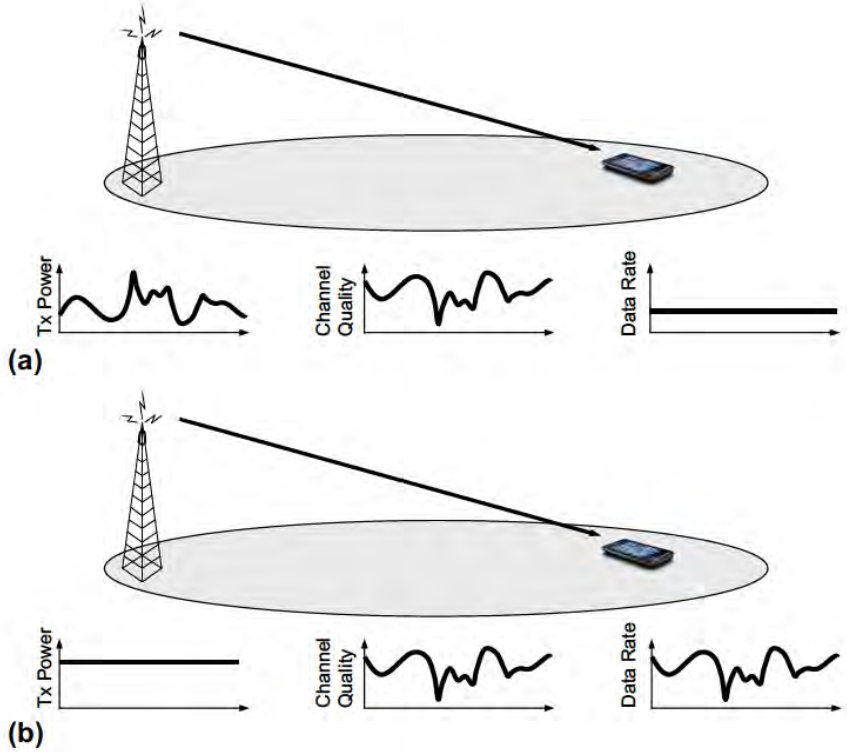


FIGURE 6.1

(a) Power control (top) and (b) rate control (bottom)

in the instantaneous channel conditions.[1] As the name suggests, dynamic power control dynamically adjusts the radio-link transmit power to compensate for variations and differences in the instantaneous channel conditions. The aim of these adjustments is to maintain a (near) constant E_b/N_0 at the receiver to successfully transmit data without the error probability becoming too high. In principle, transmit-power control increases the power at the transmitter when the radio link experiences poor radio conditions (and vice versa). Thus, the transmit power is in essence inversely proportional to the channel quality, as illustrated in Figure 6.1a. This results in a basically constant data rate, regardless of the channel variations. For services such as circuit-switched voice, this is a desirable property. Transmit-power control can be

[1]Power control is also used in GSM, but operating on a much slower basis than CDMA-based systems. In CDMA-based systems, power control is also essential to handle the near-far problem and the associated performance impact from non-orthogonal uplink transmissions.

Copyright Elsevier 2024

seen as one type of link adaptation—that is, the adjustment of transmission parameters, in this case the transmit power—to adapt to differences and variations in the instantaneous channel conditions to maintain the received E_b/N_0 at a desired level.

However, in many cases of mobile communication, especially in the case of packet-data traffic, there is not a strong need to provide a certain constant data rate over a radio link. Rather, from a user perspective, the data rate provided over the radio interface should simply be as "high as possible." Actually, even in case of typical "constant-rate" services such as voice and video (short-term), variations in the data rate are often not an issue, as long as the average data rate remains constant, assuming averaging over some relatively short time interval. In such cases—that is, when a constant data rate is not required—an alternative to transmit-power control is link adaptation by means of *dynamic rate control*. Rate control does not aim at keeping the instantaneous radio-link data rate constant, regardless of the instantaneous channel conditions. Instead, with rate control, the data rate is dynamically adjusted to compensate for the varying channel conditions. In situations with advantageous channel conditions, the data rate is increased, and vice versa. Thus, rate control maintains the $E_b/N_0 \sim P/R$ at the desired level, not by adjusting the transmission power P, but rather by adjusting the data rate R. This is illustrated in Figure 6.1b.

It can be shown that rate control is more efficient than power control [42,43]. Rate control in principle implies that the power amplifier is always transmitting at full power and therefore efficiently utilized. Power control, on the other hand, results in the power amplifier in most situations not being efficiently utilized as the transmission power is less than its maximum.

In practice, the radio-link data rate is controlled by adjusting the modulation scheme and/or the channel coding rate. In the case of advantageous radio-link conditions, the E_b/N_0 at the receiver is high and the main limitation of the data rate is the bandwidth of the radio link. Hence, as discussed in Chapter 2, in such situations higher-order modulation, for example 16QAM or 64QAM, together with a high code rate, is appropriate. Similarly, in the case of poor radio-link conditions, QPSK and low-rate coding is used. For this reason, link adaptation by means of rate control is sometimes referred to as *Adaptive Modulation and Coding* (AMC).

6.2 Channel-dependent scheduling

Scheduling controls the allocation of the shared resources among users at each time instant. It is closely related to link adaptation, and often scheduling and link adaptation are seen as one joint function. The scheduling principles, as well as which resources are shared between users, differ depending on the radio-interface characteristics—for example, whether uplink or downlink is considered and whether different users' transmissions are mutually orthogonal or not.

6.2.1 Downlink scheduling

In the downlink, transmissions to different terminals within a cell are typically mutually orthogonal, implying that, at least in theory, there is no interference between the

Copyright Elsevier 2024

transmissions (no intra-cell interference). Downlink intra-cell orthogonality can be achieved in the time domain (*Time-Division Multiplexing*, TDM), in the frequency domain (*Frequency-Domain Multiplexing*, FDM), or in the code domain (*Code-Domain Multiplexing*, CDM). In addition, the spatial domain can also be used to separate users, at least in a quasi-orthogonal way, through different antenna arrangements. This is sometimes referred to as *Spatial-Division Multiplexing* (SDM), although in most cases it is used in combination with one or several of the above multiplexing strategies.

For packet data, where the traffic is often very bursty, it can be shown that TDM is preferable from a theoretical point of view [44,45] and is therefore typically the main component in the downlink [46,47]. However, as discussed in Chapter 3, the TDM component is often combined with sharing of the radio resource also in the frequency domain (FDM) or in the code domain (CDM). For example, in the case of HSPA [96], downlink multiplexing is a combination of TDM and CDM. On the other hand, in the case of LTE, downlink multiplexing is a combination of TDM and FDM. The reasons for sharing the resources not only in the time domain will be elaborated upon later in this section.

When transmissions to multiple users occur in parallel, either by using FDM or CDM, there is also an instantaneous sharing of the total available cell transmit power. In other words, not only are the time/frequency/code resources shared resources, but also the power resource in the base station. In contrast, in the case of sharing only in the time domain there is, by definition, only a single transmission at a time and thus no instantaneous sharing of the total available cell transmit power.

For the purpose of this discussion, assume initially a TDM-based downlink with a single user being scheduled at a time. In this case, the utilization of the radio resources is maximized if, at each time instant, all resources are assigned to the user with the best instantaneous channel condition:

- In the case of link adaptation based on power control, this implies that the lowest possible transmit power can be used for a given data rate and thus minimizes the interference to transmissions in other cells for a given link utilization.
- In the case of link adaptation based on rate control, this implies that the highest data rate is achieved for a given transmit power or, in other words, for a given interference to other cells, the highest link utilization is achieved.

However, if applied to the downlink, transmit-power control in combination with TDM scheduling implies that the total available cell transmit power will, in most cases, not be fully utilized. Thus, rate control is generally preferred [42,44,47,48].

The strategy outlined above is an example of channel-dependent scheduling, where the scheduler takes the instantaneous radio-link conditions into account. Scheduling the user with the instantaneously best radio-link conditions is often referred to as *max*-C/I (or *maximum rate*) scheduling. Since the radio conditions for the different radio links within a cell typically vary independently, at each point in time there is almost always a radio link whose channel quality is near its peak (see Figure 6.2). Thus, the channel eventually used for transmission

Copyright Elsevier 2024

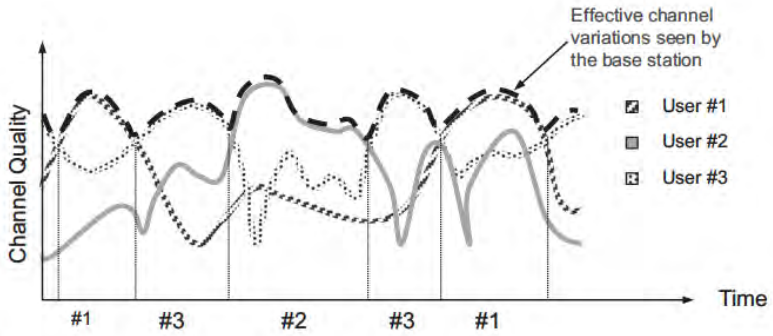


FIGURE 6.2

Channel-dependent scheduling

will typically have a high quality and, with rate control, a correspondingly high data rate can be used. This translates into a high system capacity. The gain obtained by transmitting to users with favorable radio-link conditions is commonly known as multi-user diversity; the multi-user diversity gains are larger the larger the channel variations and the larger the number of users in a cell. Hence, in contrast to the traditional view that fast fading—rapid variations in the radio-link quality—is an undesirable effect that has to be combated, the possibility of channel-dependent scheduling implies that *fading is in fact potentially beneficial and should be exploited.*

Mathematically, the max-C/I (maximum rate) scheduler can be expressed as scheduling user k given by:

$$k = \arg \max_i R_i$$

where R_i is the instantaneous data rate for user i. Although, from a system capacity perspective, max-C/I scheduling is beneficial, this scheduling principle will not be fair in all situations. If all terminals are, on average, experiencing similar channel conditions and large variations in the instantaneous channel conditions are only due to, for example, fast multi-path fading, all users will experience the same average data rate. Any variations in the instantaneous data rate are rapid and often not even noticeable by the user. However, in practice different terminals will also experience differences in the (short-term) average channel conditions—for example, due to differences in the distance and shadow fading between the base station and the terminal. In this case, the channel conditions experienced by one terminal may, for a relatively long time, be worse than the channel conditions experienced by other terminals. A pure max-C/I-scheduling strategy may then "starve" the terminal with the bad channel conditions, and the terminal with bad channel conditions will never be scheduled. This is illustrated in Figure 6.3a, where a max-C/I scheduler is used to schedule between two different users with different average channel quality. The same user

Copyright Elsevier 2024

Samsung Exhibit 1017, Page 115 of 538
Samsung Electronics Co., Ltd. v. Asus Technology Licensing Inc.
IPR2024-01217

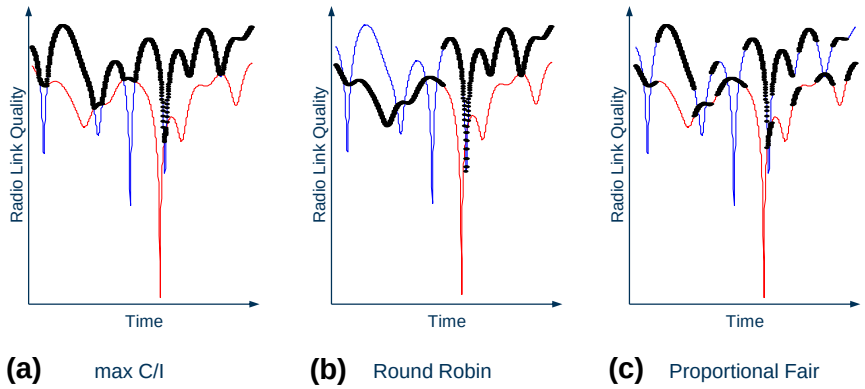


FIGURE 6.3

Example of three different scheduling behaviors for two users with different average channel quality: (a) max C/I, (b) round robin, and (c) proportional fair. The selected user is shown with bold lines

is scheduled virtually all the time. Although resulting in the highest system capacity, this situation is often not acceptable from a quality-of-service point of view.

An alternative to the max-C/I scheduling strategy is so-called *round-robin* scheduling, illustrated in Figure 6.3b. This scheduling strategy lets users take turns in using the shared resources, without taking the instantaneous channel conditions into account. Round-robin scheduling can be seen as fair scheduling in the sense that the same amount of radio resources (the same amount of time) is given to each communication link. However, round-robin scheduling is not fair in the sense of providing the same service quality to all communication links. In that case more radio resources (more time) must be given to communication links with bad channel conditions. Furthermore, as round-robin scheduling does not take the instantaneous channel conditions into account in the scheduling process, it will lead to lower overall system performance but more equal service quality between different communication links, compared to max-C/I scheduling.

Thus, what is needed is a scheduling strategy that is able to utilize the fast channel variations to improve the overall cell throughput while still ensuring the same average user throughput for all users (or at least a certain minimum user throughput for all users). When discussing and comparing different scheduling algorithms it is important to distinguish between different types of variations in the service quality:

- Fast variations in the service quality corresponding to, for example, fast multi-path fading and fast variations in the interference level. For many packet-data applications, relatively large short-term variations in service quality are often acceptable or not even noticeable to the user.

Copyright Elsevier 2024

- More long-term differences in the service quality between different communication links corresponding to, for example, differences in the distance to the cell site and shadow fading. In many cases there is a need to limit such long-term differences in service quality.

A practical scheduler should thus operate somewhere in between the max-C/I scheduler and the round-robin scheduler—that is, try to utilize fast variations in channel conditions as much as possible while still satisfying some degree of fairness between users.

One example of such a scheduler is the proportional-fair scheduler [49–51], illustrated in Figure 6.3c. In this strategy, the shared resources are assigned to the user with the *relatively* best radio-link conditions—that is, at each time instant, user k is selected for transmission according to:

$$k = \arg \max_i \frac{R_i}{\overline{R_i}}$$

where R_i is the instantaneous data rate for user i and $\overline{R_i}$ is the average data rate for user i. The average is calculated over a certain averaging period T_{PF}. To ensure efficient usage of the short-term channel variations and, at the same time, limit the long-term differences in service quality to an acceptable level, the time constant T_{PF} should be set longer than the time constant for the short-term variations. At the same time, T_{PF} should be sufficiently short so that quality variations within the interval T_{PF} are not strongly noticed by a user. Typically, T_{PF} can be set to be of the order of one second.

In the above discussion, it was assumed that all the radio resources in the downlink were assigned to a single user at a time—that is, scheduling was done purely in the time domain using TDM between users. However, in several situations, TDM is complemented by CDM or FDM. In principle, there are two reasons for not relying solely on TDM in the downlink:

- In the case of insufficient payload—that is, the amount of data to transfer to a user is not sufficiently large to utilize the full channel capacity, and a fraction of resources could be assigned to another user, either through FDM or CDM.
- In the case where channel variations in the frequency domain are exploited through FDM, as discussed further below.

The scheduling strategies in these cases can be seen as generalizations of the schemes discussed for the TDM-only cases above. For example, to handle small payloads a greedy filling approach can be used, where the scheduled user is selected according to max-C/I (or any other scheduling scheme). Once this user has been assigned resources matching the amount of data awaiting transmission, the second best user according to the scheduling strategy is selected and assigned (a fraction of) the residual resources and so on.

Finally, it should also be noted that the scheduling algorithm typically is a base-station-implementation issue and nothing that is normally specified in any standard. What needs to be specified in a standard to support channel-dependent scheduling are channel-quality measurements/reports and the signaling needed for dynamic resource allocation.

Copyright Elsevier 2024

6.2.2 Uplink scheduling

The previous section discussed scheduling from a downlink perspective. However, scheduling is equally applicable to uplink transmissions and to a large extent the same principles can be considered, although there are some differences between the two.

Fundamentally, the uplink power resource is *distributed* among the users, while in the downlink the power resource is *centralized* within the base station. Furthermore, the maximum uplink transmission power of a single terminal is typically much lower than the output power of a base station. This has a significant impact on the scheduling strategy. Unlike the downlink, where pure TDMA can often be used, uplink scheduling typically has to rely on sharing in the frequency and/or code domain in addition to the time domain, as a single terminal may not have sufficient power for efficiently utilizing the link capacity.

Similar to the downlink case, channel-dependent scheduling is also beneficial in the uplink case. However, the characteristics of the underlying radio interface, most notably whether the uplink relies on orthogonal or non-orthogonal multiple access and the type of link adaptation scheme used, also have a significant impact on the uplink scheduling strategy.

In the case of a non-orthogonal multiple-access scheme such as CDMA, power control is typically essential for proper operation. As discussed earlier in this chapter, the purpose of power control is to control the received E_b/N_0 such that the received information can be recovered. However, in a non-orthogonal multiple-access setting, power control also serves the purpose of controlling the amount of interference affecting *other* users. This can be expressed as the *maximum tolerable interference level* at the base station is a shared resource. Even if, from a single user's perspective, it would be beneficial to transmit at full power to maximize the data rate, this may not be acceptable from an interference perspective as other terminals in this case may not be able to successfully transfer any data. Thus, with non-orthogonal multiple access, scheduling a terminal when the channel conditions are favorable may not directly translate into a higher data rate as the interference generated to other simultaneously transmitting terminals in the cell must be taken into account. Stated differently, the *received* power (and thus the data rate) is, thanks to power control, in principle constant, regardless of the channel conditions at the time of transmission, while the *transmitted* power depends on the channel conditions at the time of transmission. Hence, even though channel-dependent scheduling in this example does not give a direct gain in terms of a higher data rate from the terminal, channel-dependent scheduling will still provide a gain for the system in terms of lower intra-cell interference.

The above discussion on non-orthogonal multiple access was simplified in the sense that no bounds on the terminals' transmission power were assumed. In practice, the transmission power of a terminal is upper bounded, due both to implementation and regulatory reasons, and scheduling a terminal for transmission in favorable channel conditions decreases the probability that the terminal has insufficient power to utilize the channel capacity.

In the case of an orthogonal multiple-access scheme, intra-cell power control is fundamentally not necessary and the benefits of channel-dependent scheduling become more

Copyright Elsevier 2024

Samsung Exhibit 1017, Page 118 of 538
Samsung Electronics Co., Ltd. v. Asus Technology Licensing Inc.
IPR2024-01217

similar to the downlink case. In principle, from an intra-cell perspective, a terminal can transmit at full power and the scheduler assigns a suitable part of the orthogonal resources (in practice a suitable part of the overall bandwidth) to the terminal for transmission. The remaining orthogonal resources can be assigned to other users. However, implementation constraints, for example leakage between the received signals or limited dynamic range in the receiver circuitry, may pose restrictions on the maximum tolerable power difference between the signals from simultaneously transmitting terminals. As a consequence, a certain degree of power control may be necessary, making the situation somewhat similar to the non-orthogonal case.

The discussion on non-orthogonal and orthogonal multiple access mainly considered intra-cell multiple access. However, in many practical systems universal frequency reuse between cells is applied. In this case, the inter-cell multiple access is non-orthogonal, regardless of the intra-cell multiple access, which sets limits on the allowable transmission power from a terminal.

Regardless of whether orthogonal or non-orthogonal multiple access is used, the same basic scheduling principles as for the downlink can be used. A max-C/I scheduler would assign all the uplink resources to the terminal with the best uplink channel conditions. Neglecting any power limitations in the terminal, this would result in the highest capacity (in an isolated cell) [44].

In the case of a non-orthogonal multiple-access scheme, *greedy filling* is one possible scheduling strategy [48]. With greedy filling, the terminal with the best radio conditions is assigned as high a data rate as possible. If the interference level at the receiver is smaller than the maximum tolerable level, the terminal with the second best channel conditions is allowed to transmit as well, continuing with more and more terminals until the maximum tolerable interference level at the receiver is reached. This strategy maximizes the air interface utilization but is achieved at the cost of potentially large differences in data rates between users. In the extreme case, a user at the cell border with poor channel conditions may not be allowed to transmit at all.

Strategies between greedy filling and max-C/I can also be envisioned—for example, different proportional-fair strategies. This can be achieved by including a weighting factor for each user, proportional to the ratio between the instantaneous and average data rates, into the greedy filling algorithm.

The schedulers above all assume knowledge of the instantaneous radio-link conditions, which can be hard to obtain in the uplink scenario, as discussed in Section 6.2.4. In situations when no information about the uplink radio-link quality is available at the scheduler, *round-robin* scheduling can be used. Similar to the downlink, round-robin implies terminals taking turns in transmitting, thus creating a TDMA-like operation with inter-user orthogonality in the time domain. Although the round-robin scheduler is simple, it is far from the optimal scheduling strategy.

However, as already discussed in Chapter 4, the transmission power in a terminal is limited and therefore additional sharing of the uplink resources in the frequency and/or code domain

Copyright Elsevier 2024

Samsung Exhibit 1017, Page 119 of 538
Samsung Electronics Co., Ltd. v. Asus Technology Licensing Inc.
IPR2024-01217

is required. This also impacts the scheduling decisions. For example, terminals far from the base station typically operate in the power-limited region, in contrast to terminals close to the base station, which often are in the bandwidth-limited region (for a discussion on power-limited vs. bandwidth-limited operation, see Chapter 2). Thus, for a terminal far from the base station, increasing the bandwidth will not result in an increased data rate and it is better to only assign a small amount of the bandwidth to this terminal and assign the remaining bandwidth to other terminals. On the other hand, for terminals close to the base station, an increase in the assigned bandwidth will provide a higher data rate.

6.2.3 Link adaptation and channel-dependent scheduling in the frequency domain

In the previous section, TDM-based scheduling was assumed and it was explained how, in this case, channel variations in the time domain could be utilized to improve system performance by applying channel-dependent scheduling, especially in combination with dynamic rate control. However, if the scheduler has access to the frequency domain, for example through the use of OFDM transmission, scheduling and link adaptation can also take place in the frequency domain.

Link adaptation in the frequency domain implies that, based on knowledge about the instantaneous channel conditions also in the frequency domain—that is, knowledge about the attenuation as well as the noise/interference level of, in the extreme case, every OFDM subcarrier—the power and/or the data rate of each OFDM carrier can be individually adjusted for optimal utilization.

Similarly, channel-dependent scheduling in the frequency domain implies that, based on knowledge about the instantaneous channel conditions also in the frequency domain, different subcarriers are used for transmission to or from different terminals. The scheduling gains from exploiting variations in the frequency domain are similar to those obtained from time-domain variations. In situations where the channel quality varies significantly with the frequency while the channel quality varies slowly with time, channel-dependent scheduling in the frequency domain can enhance system capacity. An example of such a situation is a wideband indoor system with low mobility, where the quality only varies slowly with time.

6.2.4 Acquiring channel-state information

To select a suitable data rate, in practice a suitable modulation scheme and channel-coding rate, the transmitter needs information about the radio-link channel conditions. Such information is also required for the purpose of channel-dependent scheduling. In the case of a system based on frequency-division duplex (FDD), only the receiver can accurately estimate the radio-link channel conditions.

For the downlink, most systems provide a downlink signal of a predetermined structure, known as the downlink pilot or the downlink reference signal. This reference signal is

Copyright Elsevier 2024

transmitted from the base station with constant power and can be used by the terminal to estimate the instantaneous downlink channel conditions. Information about the instantaneous downlink conditions can then be reported to the base station.

Basically, what is relevant for the transmitter is an estimate reflecting the channel conditions *at the time of transmission*. Hence, in principle, the terminal could apply a predictor, trying to predict the future channel conditions and report this predicted value to the base station. However, as this would require specification of prediction algorithms and how they would operate when the terminal is moving at different speeds, most practical systems simply report the measured channel conditions to the base station. This can be seen as a very simple predictor, basically assuming the conditions in the near future will be similar to the current conditions. Thus, the more rapid the time-domain channel variations, the less efficient the link adaptation.

As there inevitably will be a delay between the point in time when the terminal measured the channel conditions and the application of the reported value in the transmitter, channel-dependent scheduling and link adaptation typically operates at its best at low terminal mobility. If the terminal starts to move at high speed, the measurement reports will be outdated once arriving at the base station. In such cases, it is often preferable to perform link adaptation on the long-term average channel quality and rely on hybrid ARQ with soft combining for the rapid adaptation.

For the uplink, estimation of the uplink channel conditions is not as straightforward, as there is typically no reference signal transmitted with constant power from each terminal. Means to estimate uplink channel conditions for LTE are discussed in Chapter 13.

In the case of a system with time-division duplex (TDD), where uplink and downlink communication are time multiplexed within the same frequency band, the instantaneous uplink signal quality attenuation could be estimated from downlink measurements of the terminal, due to the reciprocity of the multi-path fading in the case of TDD. However, it should then be noted that this may not provide full knowledge of the downlink channel conditions. As an example, the interference situations at the terminal and the base station are different also in the case of TDD.

6.2.5 Traffic behavior and scheduling

It should be noted that there is little difference between different scheduling algorithms at low system load—that is, when only one or, in some cases, a few users have data waiting for transmission at the base station at each scheduling instant. The differences between different scheduling algorithms are primarily visible at high load. However, not only the load, but also the traffic behavior affects the overall scheduling performance.

As discussed above, channel-dependent scheduling tries to exploit short-term variations in radio quality. Generally, a certain degree of long-term fairness in service quality is desirable, which should be accounted for in the scheduler design. However, since system throughput decreases the more fairness is enforced, a trade-off between fairness and system throughput is

Copyright Elsevier 2024

Samsung Exhibit 1017, Page 121 of 538
Samsung Electronics Co., Ltd. v. Asus Technology Licensing Inc.
IPR2024-01217

necessary. In this trade-off, it is important to take traffic characteristics into account as they have a significant influence on the trade-off between system throughput and service quality.

To illustrate this, further consider the three different downlink schedulers discussed above:

1. *Round-robin (RR) scheduler*, where channel conditions are not taken into account.
2. *Proportional-fair (PF) scheduler*, where short-term channel variations are exploited while maintaining the long-term average user data rate.
3. *Max-C/I scheduler*, where the user with the best instantaneous channel quality in absolute terms is scheduled.

For a full buffer scenario when there is always data available at the base station for all terminals in the cell, a max-C/I scheduler will result in no, or very low, user throughput for users at the cell edge with a low average channel quality. The reason is the fundamental strategy of the max-C/I scheduler—all resources are allocated for transmission to the terminal whose channel conditions support the highest data rate. Only in the rare, not to say unlikely, case of a cell-edge user having better conditions than a cell-center user, for example due to a deep fading dip for the cell-center user, will the cell-edge user be scheduled. A proportional-fair scheduler, on the other hand, will ensure some degree of fairness by selecting the user supporting the highest data rate relative to its average data rate. Hence, users tend to be scheduled on their fading peaks, regardless of the absolute quality. Thus, users on the cell edge will also be scheduled, thereby resulting in some degree of fairness between users.

For a scenario with bursty packet data, the situation is different. In this case, the users' buffers will be finite and in many cases also empty. For example, a web page has a certain size and, after transmitting the page, there is no more data to be transmitted to the terminal in question until the user requests a new page by clicking on a link. In this case, a max-C/I scheduler can still provide a certain degree of fairness. Once the buffer for the user with the highest C/I has been emptied, another user with non-empty buffers will have the highest C/I and be scheduled, and so on. This is the reason for the difference between full buffer and web-browsing traffic illustrated in Figure 6.4. The proportional-fair scheduler has a similar performance in both scenarios.

Clearly, the degree of fairness introduced by the traffic properties depends heavily on the actual traffic; a design made with certain assumptions may be less desirable in an actual network where the traffic pattern may be different from the assumptions made during the design. Therefore, relying solely on the traffic properties for fairness is not a good strategy, but the discussion above emphasizes the need to design the scheduler not only for the full buffer case.

6.3 Advanced retransmission schemes

Transmissions over wireless channels are subject to errors, for example due to variations in the received signal quality. To some degree, such variations can be counteracted through link

Copyright Elsevier 2024

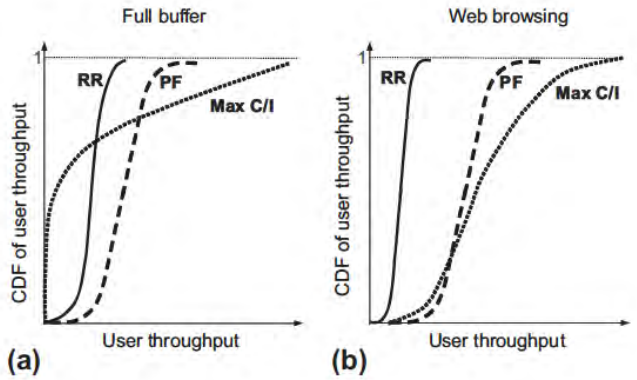


FIGURE 6.4

Illustration of the principle behavior of different scheduling strategies: (a) for full buffers and (b) for web browsing traffic model

adaptation, as discussed above. However, receiver noise and unpredictable interference variations cannot be counteracted. Therefore, virtually all wireless communications systems employ some form of *Forward Error Correction* (FEC), tracing its roots to the pioneering work of Claude Shannon in 1948 [11]. There is a rich literature in the area of error-correction coding (see, for example, [52,53] and references therein), and a detailed description is beyond the scope of this book. In short, the basic principle behind forward error-correction coding is to introduce redundancy in the transmitted signal. This is achieved by adding *parity bits* to the information bits prior to transmission (alternatively, the transmission could consist of parity bits alone, depending on the coding scheme used). The parity bits are computed from the information bits using a method given by the coding structure used. Thus, the number of bits transmitted over the channel is larger than the number of original information bits and a certain amount of *redundancy* has been introduced in the transmitted signal.

Another approach to handle transmissions errors is to use *Automatic Repeat Request* (ARQ). In an ARQ scheme, the receiver uses an error-detection code, typically a Cyclic Redundancy Check (CRC), to detect if the received packet is in error or not. If no error is detected in the received data packet, the received data is declared error-free and the transmitter is notified by sending a positive acknowledgement (ACK). On the other hand, if an error is detected, the receiver discards the received data and notifies the transmitter via a return channel by sending a negative acknowledgement (NAK). In response to an NAK, the transmitter retransmits the same information.

Virtually all modern communication systems, including LTE, employ a combination of forward error-correction coding and ARQ, known as *hybrid ARQ*. Hybrid ARQ uses forward error-correction codes to correct a subset of all errors and relies on error detection to detect uncorrectable errors. Erroneously received packets are discarded and the receiver requests

Copyright Elsevier 2024

Samsung Exhibit 1017, Page 123 of 538
Samsung Electronics Co., Ltd. v. Asus Technology Licensing Inc.
IPR2024-01217

retransmissions of corrupted packets. Thus, it is a combination of FEC and ARQ, as described above. Hybrid ARQ was first proposed in [54] and numerous publications on hybrid ARQ have appeared since (see [52] and references therein). Most practical hybrid ARQ schemes are built around a CRC code for error detection and convolutional or Turbo codes for error correction, but in principle any error-detection and error-correction code can be used.

6.4 Hybrid ARQ with soft combining

The hybrid-ARQ operation described above discards erroneously received packets and requests retransmission. However, despite it not being possible to decode the packet, the received signal still contains information, which is lost by discarding erroneously received packets. This shortcoming is addressed by *hybrid ARQ with soft combining*. In hybrid ARQ with soft combining, the erroneously received packet is stored in a buffer memory and later combined with the retransmission to obtain a single, combined packet that is more reliable than its constituents. Decoding of the error-correction code operates on the combined signal. If the decoding fails (typically a CRC code is used to detect this event), a retransmission is requested.

Retransmission in any hybrid-ARQ scheme must, by definition, represent the same set of information bits as the original transmission. However, the set of coded bits transmitted in each retransmission may be selected differently as long as they represent the same set of information bits. Hybrid ARQ with soft combining is therefore usually categorized into *Chase combining* and *incremental redundancy*, depending on whether the retransmitted bits are required to be identical to the original transmission or not.

Chase combining, where the retransmissions consist of the same set of coded bits as the original transmission, was first proposed in [55]. After each retransmission, the receiver uses maximum-ratio combining to combine each received channel bit with any previous transmissions of the same bit, and the combined signal is fed to the decoder. As each retransmission is an identical copy of the original transmission, retransmissions with Chase combining can be seen as additional repetition coding. Therefore, as no new redundancy is transmitted, Chase combining does not give any additional coding gain but only increases the accumulated received E_b/N_0 for each retransmission (Figure 6.5).

Several variants of Chase combining exist. For example, only a subset of the bits transmitted in the original transmission might be retransmitted, so-called partial Chase combining. Furthermore, although combining is often done after demodulation but before channel decoding, combining can also be carried out at the modulation symbol level before demodulation, as long as the modulation scheme is unchanged between transmission and retransmission.

With *Incremental Redundancy* (IR), each retransmission does not have to be identical to the original transmission. Instead, *multiple sets* of coded bits are generated, each representing the same set of information bits [56,57]. Whenever a retransmission is required, the

Copyright Elsevier 2024

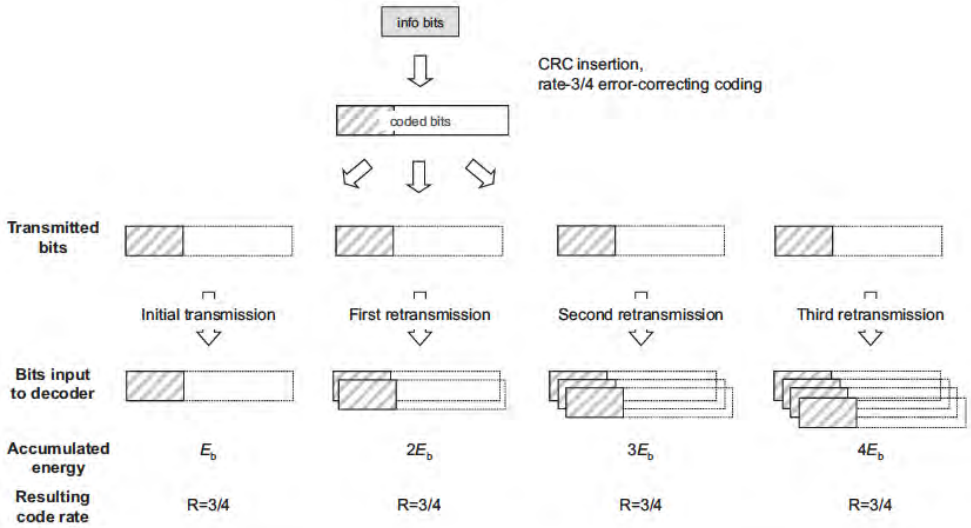


FIGURE 6.5

Example of Chase combining

retransmission typically uses a different set of coded bits than the previous transmission. The receiver combines the retransmission with previous transmission attempts of the same packet. As the retransmission may contain additional parity bits not included in the previous transmission attempts, the resulting code rate is generally lowered by a retransmission. Furthermore, each retransmission does not necessarily have to consist of the same number of coded bits as the original and, in general, the modulation scheme can also be different for different retransmissions. Hence, incremental redundancy can be seen as a generalization of Chase combining or, stated differently, Chase combining is a special case of incremental redundancy.

Typically, incremental redundancy is based on a low-rate code and the different redundancy versions are generated by puncturing the output of the encoder. In the first transmission only a limited number of coded bits are transmitted, effectively leading to a high-rate code. In the retransmissions, additional coded bits are transmitted. As an example, assume a basic rate-1/4 code. In the first transmission, only every third coded bit is transmitted, effectively giving a rate-3/4 code, as illustrated in Figure 6.6. In the case of a decoding error and a subsequent request for a retransmission, additional bits are transmitted, effectively leading to a rate-3/8 code. After a second retransmission the code rate is 1/4. In the case of more than two retransmissions, transmitted coded bits would already be repeated. In addition to a gain in accumulated received E_b/N_0, incremental redundancy also results in a coding gain for each

Copyright Elsevier 2024

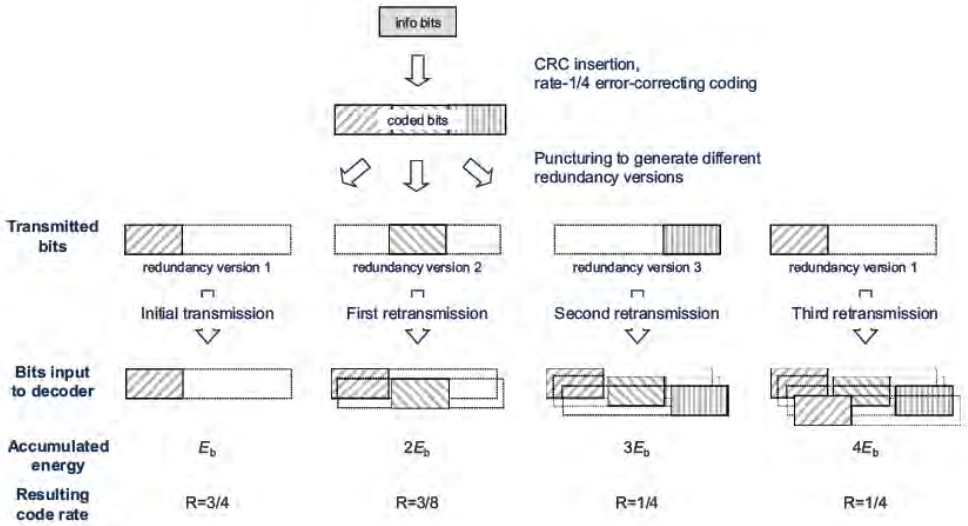


FIGURE 6.6

Example of incremental redundancy

retransmission. The gain with IR compared to Chase is larger for high initial code rates, while at lower initial coding rates Chase combining is almost as good as IR [58]. Furthermore, as shown in [59], the performance gain of incremental redundancy compared to Chase combining can also depend on the relative power difference between the transmission attempts.

With incremental redundancy, the code used for the first transmission should provide good performance not only when used alone, but also when used in combination with the code for the second transmission. The same holds for subsequent retransmissions. Thus, as the different redundancy versions typically are generated through puncturing of a low-rate mother code, the puncturing patterns should be defined such that all the code bits used by a high-rate code should also be part of any lower-rate codes. In other words, the resulting code rate R_i after transmission attempt i, consisting of the coded bits from redundancy versions RV_k, $k = 1, \ldots, i$, should have similar performance as a good code designed directly for rate R_i. Examples of this for convolutional codes are the so-called rate-compatible convolutional codes [60].

In the discussion so far, it has been assumed that the receiver has received all the previously transmitted redundancy versions. If all redundancy versions provide the same amount of information about the data packet, the order of the redundancy versions is not critical. However, for some code structures, not all redundancy versions are of equal importance. One example here is Turbo codes, where the systematic bits are of higher importance than the

Copyright Elsevier 2024

parity bits. Hence, the initial transmission should at least include all the systematic bits and some parity bits. In the retransmission(s), parity bits not in the initial transmission can be included. However, if the initial transmission was received with poor quality or not at all, a retransmission with only parity bits is not appropriate. A retransmission of (some of) the systematic bits provides better performance. Incremental redundancy with Turbo codes can therefore benefit from multiple levels of feedback, for example by using two different negative acknowledgements—NAK to request additional parity bits, and LOST to request a retransmission of the systematic bits. In general, the problem of determining the amount of systematic and parity bits in a retransmission based on the signal quality of previous transmission attempts is non-trivial.

Hybrid ARQ with soft combining, regardless of whether Chase or incremental redundancy is used, leads to an implicit reduction of the data rate by means of retransmissions and can thus be seen as implicit link adaptation. However, in contrast to link adaptation based on explicit estimates of the instantaneous channel conditions, hybrid ARQ with soft combining implicitly adjusts the coding rate based on the result of the decoding. In terms of overall throughput this kind of implicit link adaptation can be superior to explicit link adaptation, as additional redundancy is only added *when needed*—that is, when it is not possible to correctly decode previous higher-rate transmissions. Furthermore, as hybrid ARQ does not try to predict any channel variations, it works equally well, regardless of the speed at which the terminal is moving. Since implicit link adaptation can provide a gain in system throughput, a valid question is why explicit link adaptation is necessary at all. One major reason for having explicit link adaptation is the reduced delay. Although relying on implicit link adaptation alone is sufficient from a system throughput perspective, the end-user service quality may not be acceptable from a delay perspective.

Copyright Elsevier 2024

This page intentionally left blank

Copyright Elsevier 202

CHAPTER

LTE Radio Access: An Overview

7

The work on LTE was initiated in late 2004 with the overall aim of providing a new radio-access technology focusing on packet-switched data only. The first phase of the 3GPP work on LTE was to define a set of performance and capability targets for LTE [10]. This included targets on peak data rates, user/system throughput, spectral efficiency, and control/user-plane latency. In addition, requirements were also set on spectrum flexibility, as well as on interaction/compatibility with other 3GPP radio-access technologies (GSM, WCDMA/HSPA, and TD-SCDMA).

Once the targets were set, 3GPP studies on the feasibility of different technical solutions considered for LTE were followed by development of detailed specifications. The first release of the LTE specifications, release 8, was completed in the spring of 2008 and commercial network operation began in late 2009. Release 8 has been followed by additional LTE releases, introducing more functionality and capabilities in different areas, as illustrated in Figure 7.1. LTE Release 11 was finalized late 2012, and 3GPP is as of this writing working on LTE release 12.

In parallel to the development of LTE, there has also been an evolution of the overall 3GPP network architecture, termed *System Architecture Evolution* (SAE), including both the radio-access network and the core network. Requirements were also set on the architecture evolution, leading to a new flat radio-access-network architecture with a single type of node, the *eNodeB*,[1] as well as a new core-network architecture. An excellent description of the LTE-associated core-network architecture, the *Evolved Packet Core* (EPC), can be found in [9].

The remaining part of this chapter provides an overview of LTE. The most important technology components of LTE release 8—including transmission schemes, scheduling, multi-antenna support, and spectrum flexibility—are presented, as well as the additional features and enhancements introduced in later releases up to and including release 11. The chapter can either be read on its own to get a high-level overview of LTE, or as an introduction to the subsequent chapters.

The following chapters, Chapters 8–20, then provide a detailed description of the LTE radio-access technology. Chapter 8 provides an overview of the LTE protocol structure, including RLC, MAC, and the physical layer, explaining the logical and physical channels and the related data flow. The time–frequency structure on which LTE is based is covered in

[1]eNodeB is a 3GPP term that can roughly be seen as being equivalent to a base station (see further Chapter 8).

4G: LTE/LTE-Advanced for Mobile Broadband. http://dx.doi.org/10.1016/B978-0-12-419985-9.00007-6
Copyright © 2014 Erik Dahlman, Stefan Parkvall and Johan Sköld. Published by Elsevier Ltd. All rights reserved.

Copyright Elsevier 2024

Samsung Exhibit 1017, Page 129 of 538
Samsung Electronics Co., Ltd. v. Asus Technology Licensing Inc.
IPR2024-01217

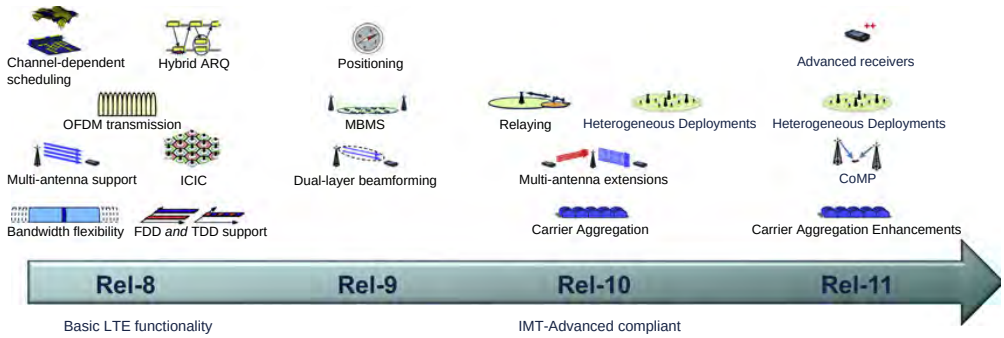


FIGURE 7.1

LTE and its evolution

Chapter 9, together with a brief overview of the LTE antenna-port concept. This is followed by a detailed description of the LTE physical-layer functionality for downlink and uplink transmission in Chapters 10 and 11 respectively. Chapter 12 contains a description of LTE retransmission mechanisms, followed by a discussion on the mechanisms available in LTE to support advanced scheduling and link adaptation in Chapter 13. Access procedures, necessary for a terminal to connect to the network, are the topic of Chapter 14. Multi-point coordination/transmission is discussed in Chapter 15, followed by a discussion of heterogeneous deployments based on LTE in Chapter 16. Chapter 17 covers the multicast/broadcast functionality of LTE, followed by a description of relaying operation in Chapter 18. Chapter 19 addresses how radio-frequency (RF) requirements are defined in LTE, taking into account the spectrum flexibility. An assessment of the system performance of LTE is found in Chapter 20. Finally, Chapter 21 concludes with a summary of the most important features expected to be introduced as part of LTE release 12, as well as a brief discussion on key challenges and some possible technology components for future beyond-LTE radio access.

7.1 Basic principles

Building upon the basic technologies described in the previous chapters, the main principles behind LTE will be described in the following sections.

7.1.1 Transmission scheme

The LTE downlink transmission scheme is based on conventional OFDM. As discussed in Chapter 3, OFDM is an attractive transmission scheme for several reasons. Due to the relatively long OFDM symbol time in combination with a cyclic prefix, OFDM provides

Copyright Elsevier 2024

a high degree of robustness against channel frequency selectivity. Although signal corruption due to a frequency-selective channel can, in principle, be handled by equalization at the receiver side, the complexity of such equalization starts to become unattractively high for implementation in a terminal at larger bandwidths and especially in combination with advanced multi-antenna transmission schemes such as spatial multiplexing. Therefore, OFDM is an attractive choice for LTE for which a wide bandwidth and support for advanced multi-antenna transmission were key requirements.

OFDM also provides some additional benefits relevant for LTE:

- OFDM provides access to the frequency domain, thereby enabling an additional degree of freedom to the channel-dependent scheduler compared to time-domain-only scheduling used in major 3G systems.
- Flexible transmission bandwidth to support operation in spectrum allocations of different sizes is straightforward with OFDM, at least from a baseband perspective, by varying the number of OFDM subcarriers used for transmission. Note, however, that support of a flexible transmission bandwidth also requires flexible RF filtering, etc., for which the exact transmission scheme is to a large extent irrelevant. Nevertheless, maintaining the same baseband-processing structure, regardless of the bandwidth, eases terminal development and implementation.
- Broadcast/multicast transmission, where the same information is transmitted from multiple base stations, is straightforward with OFDM as described in Chapter 3 and elaborated upon from an LTE perspective in Chapter 17.

The LTE uplink is also based on OFDM transmission. However, different means are taken to reduce the *cubic metric* (see Chapter 4) of uplink transmissions, thereby enabling higher power-amplifier efficiency on the terminal side. More specifically, for uplink data transmission the OFDM modulator is preceded by a DFT precoder, leading to *DFT-spread OFDM* (DFTS-OFDM) as described in Chapter 4. Often the term DFTS-OFDM is used to describe the LTE uplink transmission scheme in general. However, it should be understood that DFTS-OFDM is only used for uplink data transmission. As described in more detail in later chapters, other means are used to achieve a low cubic metric for other types of uplink transmissions. Thus, the LTE uplink transmission scheme should be described as OFDM with different techniques, including DFT precoding for data transmission, being used to reduce the cubic metric of the transmitted signal.

The use of OFDM transmission for the LTE uplink allows for orthogonal separation of uplink transmissions also in the frequency domain. Orthogonal separation is in many cases beneficial as it avoids interference between uplink transmissions from different terminals within the cell (*intra-cell interference*). As discussed in Chapter 4, allocating a very large instantaneous bandwidth for transmission from a single terminal is not an efficient strategy in situations where the data rate is mainly limited by the available terminal transmit power rather than the bandwidth. In such situations a terminal can instead be allocated only a part of the total available bandwidth and other terminals within the cell can be scheduled to transmit in

Copyright Elsevier 2024

Samsung Exhibit 1017, Page 131 of 538
Samsung Electronics Co., Ltd. v. Asus Technology Licensing Inc.
IPR2024-01217

parallel on the remaining part of the spectrum. In other words, the LTE uplink transmission scheme allows for both *time division* (TDMA) and *frequency division* (FDMA) between users.

7.1.2 Channel-dependent scheduling and rate adaptation

At the core of the LTE transmission scheme is the use of *shared-channel transmission* with the overall time–frequency resource dynamically shared between users. The use of shared-channel transmission is well matched to the rapidly varying resource requirements posed by packet-data communication and also enables several of the other key technologies on which LTE is based.

The scheduler controls, for each time instant, to which users the different parts of the shared resource should be assigned. The scheduler also determines the data rate to be used for each transmission. Thus, *rate adaptation* can be seen as a part of the scheduling functionality. The scheduler is a key element and to a large extent determines the overall system performance, especially in a highly loaded network. Both downlink and uplink transmissions are subject to tight scheduling in LTE. From Chapter 6 it is known that a substantial gain in system capacity can be achieved if the channel conditions are taken into account in the scheduling decision, so-called *channel-dependent scheduling*. Due to the use of OFDM in both the downlink and uplink transmission directions, the scheduler has access to both the time and frequency domains. In other words, the scheduler can, for each time instant and frequency region, select the user with the best channel conditions, as illustrated in Figure 7.2.

The possibility of channel-dependent scheduling in the frequency domain is particularly useful at low terminal speeds—in other words, when the channel is varying slowly in time. As discussed in Chapter 6, channel-dependent scheduling relies on channel-quality variations between users to obtain a gain in system capacity. For delay-sensitive services, a time-domain-only scheduler may, due to the delay constraints, be forced to schedule a particular user, despite the channel quality not being at its peak. In such situations, exploiting channel-quality variations also in the frequency domain will help to improve the overall performance of the system. For LTE, scheduling decisions can be taken as often as once every 1 ms and the granularity in the frequency domain is 180 kHz. This allows for relatively rapid channel variations in both the time and frequency domains to be tracked and utilized by the scheduler.

To support downlink scheduling, a terminal may provide the network with *channel-state* reports indicating the instantaneous downlink channel quality in both the time and frequency domains. The channel state is typically obtained by measuring on *reference signals* transmitted in the downlink. Based on the channel-state reports, also referred to as *channel-state information* (CSI), the downlink scheduler can assign resources for downlink transmission to different terminals, taking the channel quality into account in the scheduling decision. In principle, a scheduled terminal can be assigned an arbitrary combination of 180 kHz wide *resource blocks* in each 1 ms scheduling interval.

As already mentioned, the LTE uplink is based on orthogonal separation of different uplink transmissions and it is the task of the uplink scheduler to assign resources in both the

Copyright Elsevier 2024

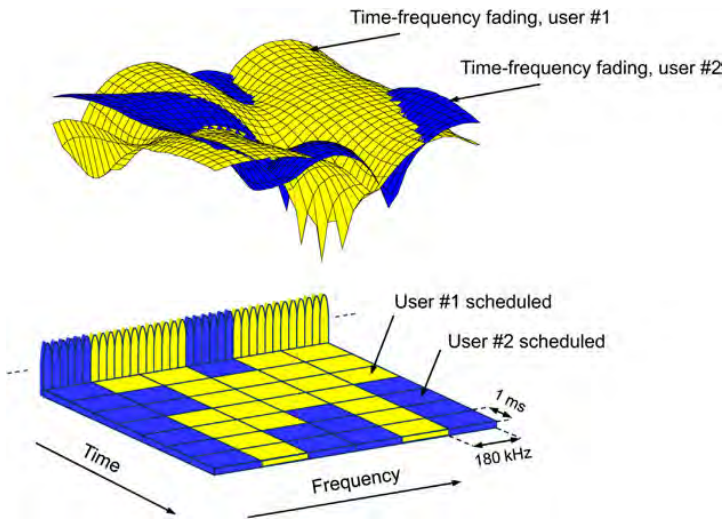


FIGURE 7.2

Downlink channel-dependent scheduling in time and frequency domains

time and frequency domains to different terminals. Scheduling decisions, taken once per 1 ms, control what set of terminals are allowed to transmit within a cell during a given time interval and, for each terminal, on what frequency resources the transmission is to take place, and what transmission parameters, including the data rate, to use.

Channel conditions can also be taken into account in the uplink scheduling process, similar to the downlink scheduling. However, as will be discussed in more detail in subsequent chapters, obtaining information about the uplink channel conditions may not be feasible or desirable in all situations. Therefore, different means to obtain *uplink diversity* are important as a complement in situations where uplink channel-dependent scheduling is not suitable.

7.1.3 Inter-cell interference coordination

LTE is designed to operate with a one-cell frequency reuse, implying that the same carrier frequency can be used at neighboring transmission points. In particular, the basic control channels of LTE are designed to operate properly with the relatively low signal-to-interference ratio that may be experienced in a reuse-one deployment.

Fundamentally, having access to all available frequency resources at each transmission point is always beneficial. However, system efficiency as well as end-user quality can be further improved if transmissions from neighboring transmission points are coordinated in such a way that the most severe interference situations can be avoided.

Copyright Elsevier 2024

Already the first release of LTE included explicit support for such coordination, in the release-8 context referred to as *Inter-cell Interference Coordination* (ICIC). More specifically, the release-8 specifications defined a set of messages that can be exchanged between eNodeBs using the so-called *X2 interface* (see Chapter 8). These messages provide information about the interference situation experienced by the eNodeB issuing the message and can be used by a neighboring eNodeB receiving the message as input to its scheduling process, thereby providing a means for at least partly coordinating the transmissions and controlling the interference between cells of different eNodeBs. The specification support for coordination between transmission points was then extended as part of the LTE release-11 "CoMP" activities, as discussed further in Section 7.4.1. Chapter 15 provides a more in-depth discussion of coordination between transmission points for LTE, covering both the release-8 ICIC mechanisms and the release-11 CoMP features.

Especially severe interference situations may occur in so-called *heterogeneous network deployments* consisting of overlapping layers of base stations with large differences in the downlink transmission power. This will be briefly discussed in Section 7.3.4 as part of the discussion on LTE release-10 features. The interference handling in heterogeneous network deployments will then be discussed in more detail in Chapter 16.

7.1.4 Hybrid ARQ with soft combining

Fast hybrid ARQ with soft combining is used in LTE to allow the terminal to rapidly request retransmissions of erroneously received transport blocks and to provide a tool for implicit rate adaptation. Retransmissions can be rapidly requested after each packet transmission, thereby minimizing the impact on end-user performance from erroneously received packets. Incremental redundancy is used as the soft combining strategy and the receiver buffers the soft bits to be able to perform soft combining between transmission attempts.

7.1.5 Multi-antenna transmission

Already from its first release, LTE has included support for different multi-antenna techniques as an integral part of the radio-interface specifications. In many respects, the use of multiple antennas is the key technology to reach many of the aggressive LTE performance targets. As discussed in Chapter 5, multiple antennas can be used in different ways for different purposes:

- Multiple receive antennas can be used for receive diversity. For uplink transmissions, they have been used in many cellular systems for several years. However, as dual receive antennas are the baseline for all LTE terminals, the downlink performance is also improved. The simplest way of using multiple receive antennas is classical receive diversity to collect additional energy and suppress fading, but additional gains can be achieved in interference-limited scenarios if the antennas are used not only to provide diversity, but also to suppress interference, as discussed in Chapter 5.

Copyright Elsevier 2024

- Multiple transmit antennas at the base station can be used for transmit diversity and different types of beam-forming. The main goal of beam-forming is to improve the received SINR and, eventually, improve system capacity and coverage.
- *Spatial multiplexing* (sometimes referred to as MIMO or more specifically SU-MIMO) using multiple antennas at both the transmitter and receiver is supported by LTE. Spatial multiplexing results in an increased data rate, channel conditions permitting, in bandwidth-limited scenarios by creating several parallel "channels," as described in Chapter 5. Alternatively, by combining the spatial properties with the appropriate interference-suppressing receiver processing, multiple terminals can transmit on the same time–frequency resource in order to improve the overall cell capacity. In 3GPP this is referred to as *multi-user MIMO*.

In general, the different multi-antenna techniques are beneficial in different scenarios. As an example, at relatively low SINR, such as at high load or at the cell edge, spatial multiplexing provides relatively limited benefits. Instead, in such scenarios multiple antennas at the transmitter side should be used to raise the SINR by means of beam-forming. On the other hand, in scenarios where there already is a relatively high SINR, for example in small cells, raising the signal quality further provides relatively minor gains as the achievable data rates are then mainly bandwidth limited rather than SINR limited. In such scenarios, spatial multiplexing should instead be used to fully exploit the good channel conditions. The multi-antenna scheme used is under control of the base station, which therefore can select a suitable scheme for each transmission.

Up to four layers of downlink spatial multiplexing were already supported from the first release of LTE. Later releases further extended the LTE multi-antenna capabilities as described in Section 7.3.2.

7.1.6 Spectrum flexibility

A high degree of spectrum flexibility is one of the main characteristics of the LTE radio-access technology. The aim of this spectrum flexibility is to allow for the deployment of LTE radio access in different frequency bands with various characteristics, including different duplex arrangements and different sizes of the available spectrum. Chapter 19 outlines further details of how spectrum flexibility is achieved in LTE.

7.1.6.1 *Flexibility in duplex arrangements*

One important part of the LTE requirements in terms of spectrum flexibility is the possibility to deploy LTE-based radio access in both paired *and* unpaired spectrum. Therefore, LTE supports both frequency- and time-division-based duplex arrangements. *Frequency-Division Duplex* (FDD), as illustrated on the left in Figure 7.3, implies that downlink and uplink transmission take place in different, sufficiently separated, frequency bands. *Time-Division Duplex* (TDD), as illustrated on the right in Figure 7.3, implies that downlink and uplink transmission take place in different, non-overlapping time slots. Thus, TDD can operate in

Copyright Elsevier 2024

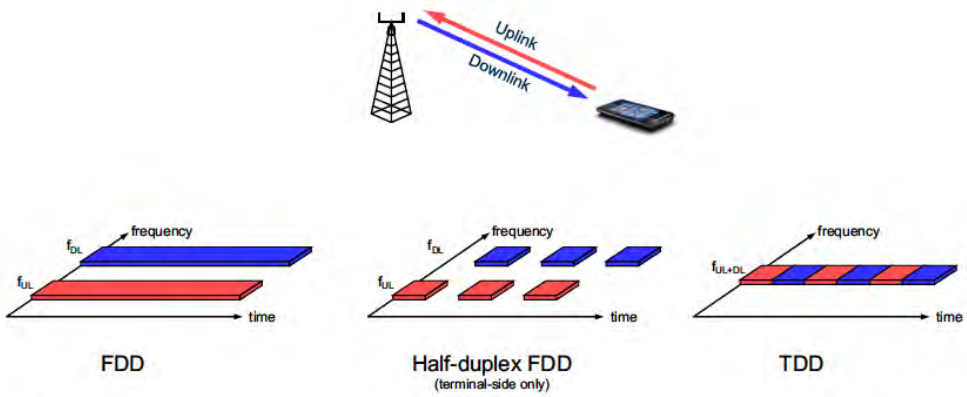


FIGURE 7.3

Frequency- and time-division duplex

unpaired spectrum, whereas FDD requires paired spectrum. The required flexibility and resulting requirements to support LTE operation in different paired and unpaired frequency arrangements are further discussed in Chapter 19.

Operation in both paired and unpaired spectrum has been supported by 3GPP radio-access technologies even before the introduction of LTE by means of FDD-based WCDMA/HSPA in combination with TDD-based TD-SCDMA radio. However, this was then achieved by means of, at least in the details, relatively different radio-access technologies leading to additional effort and complexity when developing and implementing dual-mode terminals capable of both FDD and TDD operation. LTE, on the other hand, supports both FDD and TDD *within a single radio-access technology*, leading to a minimum of deviation between FDD and TDD for LTE-based radio access. As a consequence of this, the overview of the LTE radio access provided in the following chapters is, to a large extent, valid for both FDD and TDD. In the case of differences between FDD and TDD, these differences will be explicitly indicated. Furthermore, the TDD mode, also known as TD-LTE, is designed with coexistence between LTE (TDD) and TD-SCDMA in mind to simplify a gradual migration from TD-SCDMA to LTE.

LTE also supports *half-duplex* FDD at the terminal (illustrated in the middle of Figure 7.3). In half-duplex FDD, transmission and reception *at a specific terminal* are separated in both frequency and time. The base station still uses full-duplex FDD as it simultaneously may schedule *different* terminals in uplink and downlink; this is similar to, for example, GSM operation. The main benefit with half-duplex FDD is the reduced terminal complexity as no duplex filter is needed in the terminal. This is especially beneficial in the case of multi-band terminals which otherwise would need multiple sets of duplex filters.

Copyright Elsevier 2024

7.1.6.2 Bandwidth flexibility

An important characteristic of LTE is the support for a range of different transmission bandwidths on both downlink and uplink. The main reason for this is that the amount of spectrum available for LTE deployment may vary significantly between different frequency bands and also depending on the exact situation of the operator. Furthermore, the possibility of operating in different spectrum allocations gives the possibility for gradual migration of spectrum from other radio-access technologies to LTE.

LTE supports operation in a wide range of spectrum allocations, achieved by a flexible transmission bandwidth being part of the LTE specifications. To efficiently support very high data rates when spectrum is available, a wide transmission bandwidth is necessary, as discussed in Chapter 2. However, a sufficiently large amount of spectrum may not always be available, either due to the band of operation or due to a gradual migration from another radio-access technology, in which case LTE can be operated with a more narrow transmission bandwidth. In such cases, the maximum achievable data rates will be reduced accordingly. As discussed below, the spectrum flexibility is further improved in later releases of LTE.

The LTE physical-layer specifications [61–64] are bandwidth agnostic and do not make any particular assumption on the supported transmission bandwidths beyond a minimum value. As will be seen in the following, the basic radio-access specification, including the physical-layer and protocol specifications, allows for any transmission bandwidth ranging from roughly 1 MHz up to around 20 MHz. At the same time, at an initial stage, radio-frequency requirements are only specified for a limited subset of transmission bandwidths, corresponding to what is predicted to be relevant spectrum-allocation sizes and relevant migration scenarios. Thus, in practice the LTE radio-access technology supports a limited set of transmission bandwidths, but additional transmission bandwidths can easily be introduced by updating only the RF specifications.

7.2 LTE release 9

After completing the first release of LTE, work continued in 3GPP with introducing additional functionality in the second release of the LTE specifications, release 9. The main new features introduced as part of release 9, completed in late 2009, were support for multicast transmission, support for network-assisted positioning services, and enhancements to beamforming in the downlink.

7.2.1 Multicast and broadcast support

Multi-cell broadcast implies transmission of the same information from multiple cells, as described in Chapter 3. By exploiting this at the terminal, effectively using signal power from multiple cell sites at the detection, a substantial improvement in coverage can be achieved. By transmitting not only identical signals from multiple cell sites (with

Copyright Elsevier 2024

identical coding and modulation), but also synchronizing the transmission timing between the cells, the signal at the terminal will appear exactly as a signal transmitted from a single cell site and subject to multi-path propagation. Due to the OFDM robustness to multi-path propagation, such multi-cell transmission, in 3GPP also referred to as *Multicast/Broadcast Single-Frequency Network* (MBSFN) transmission, will then not only improve the received signal strength, but also eliminate the inter-cell interference as described in Chapter 3. Thus, with OFDM, multi-cell broadcast/multicast throughput may eventually be limited by noise only and can then, in the case of small cells, reach extremely high values.

It should be noted that the use of MBSFN transmission for multi-cell broadcast/multicast assumes the use of tight synchronization and time alignment of the signals transmitted from different cell sites.

7.2.2 Positioning

Positioning, as the name implies, refers to functionality in the radio-access network to determine the location of individual terminals. Determining the position of a terminal can, in principle, be done by including a GPS receiver in the terminal. Although this is a quite common feature, not all terminals include the necessary GPS receiver and there may also be cases when the GPS service is not available. LTE release 9 therefore introduced positioning support inherent in the radio-access network. By measuring on special reference signals transmitted regularly from different cell sites, the location of the terminal can be determined.

7.2.3 Dual-layer beam-forming

Release 9 enhanced the support for combining spatial multiplexing with beam-forming. Although the combination of beam-forming and spatial multiplexing was already possible in release 8, it was restricted to so-called *codebook-based precoding* (see Chapter 10). In release 9, the support for spatial multiplexing combined with so-called *non-codebook-based precoding* was introduced, allowing for improved flexibility in deploying various multi-antenna transmission schemes.

7.3 LTE release 10 and IMT-advanced

As described in Chapter 1, *IMT-Advanced* is the term used by the ITU for radio-access technologies beyond IMT-2000. As a first step in defining IMT-Advanced, the ITU defined a set of requirements that any IMT-Advanced compliant technology should fulfill [65]. Examples of these requirements were support for at least 40 MHz bandwidth, peak spectral efficiencies of 15 bit/s/Hz in downlink and 6.75 bit/s/Hz in uplink (corresponding to peak

Copyright Elsevier 2024

rates of at least 600 and 270 Mbit/s respectively), and control and user plane latency of less than 100 and 10 ms respectively.

One of the main targets of LTE release 10 was to ensure that the LTE radio-access technology would be fully compliant with the IMT-Advanced requirements, thus the name *LTE-Advanced* is often used for LTE release 10 and later. However, in addition to the ITU requirements, 3GPP also defined its own targets and requirements for LTE-Advanced (LTE release 10) [66]. These targets/requirements extended the ITU requirements both in terms of being more aggressive as well as including additional requirements. One important requirement was *backwards compatibility*. Essentially this means that an earlier-release LTE terminal should be able to access a carrier supporting LTE release-10 functionality, although not being able to utilize all the release-10 features of that carrier.

LTE release 10 was completed in late 2010 and implied enhanced LTE spectrum flexibility through *carrier aggregation*, further extended multi-antenna transmission, introduced support for *relaying*, and provided improvements in the area of inter-cell interference coordination in *heterogeneous network deployments*.

7.3.1 Carrier aggregation

As mentioned earlier, the first release of LTE already provided extensive support for deployment in spectrum allocations of various characteristics, with bandwidths ranging from roughly 1 MHz up to 20 MHz in both paired and unpaired bands. With LTE release 10 the transmission bandwidth can be further extended by means of so-called *carrier aggregation* (CA), where multiple *component carriers* are aggregated and jointly used for transmission to/from a single terminal. Up to five component carriers, possibly each of different bandwidths, can be aggregated, allowing for transmission bandwidths up to 100 MHz. Backwards compatibility was ensured as each component carrier uses the release-8 structure. Hence, to a release-8/9 terminal each component carrier will appear as an LTE release-8 carrier, while a carrier-aggregation-capable terminal can exploit the total aggregated bandwidth, enabling higher data rates. In the general case, a different number of component carriers can be aggregated for the downlink and uplink.

Component carriers do not have to be contiguous in frequency, which enables exploitation of *fragmented spectrum*; operators with a fragmented spectrum can provide high-data-rate services based on the availability of a wide overall bandwidth even though they do not possess a single wideband spectrum allocation.

From a baseband perspective, there is no difference between the cases in Figure 7.4 and they are all supported by LTE release 10. However, the RF-implementation complexity is vastly different with the first case being the least complex. Thus, although carrier aggregation is supported by the basic specifications, not all terminals will support it. Furthermore, release 10 has some restrictions on carrier aggregation in the RF specification, compared to what has been specified for physical layer and signaling, while in release 11 there is support for carrier-aggregation within and between a much larger number of bands.

Copyright Elsevier 2024

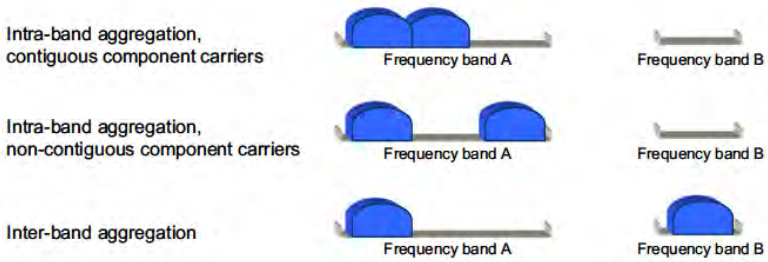


FIGURE 7.4

Carrier aggregation

7.3.2 Extended multi-antenna transmission

In release 10, downlink spatial multiplexing was expanded to support up to eight transmission layers. This can be seen as an extension of the release-9 dual-layer beam-forming to support up to eight antenna ports and eight corresponding layers. Together with the support for carrier aggregation this enables downlink data rates up to 3 Gbit/s in 100 MHz of spectrum.

Uplink spatial multiplexing of up to four layers was also introduced as part of LTE release 10. Together with the possibility for uplink carrier aggregations, this allows for uplink data rates up to 1.5 Gbit/s in 100 MHz of spectrum. Uplink spatial multiplexing consists of a codebook-based scheme under control of the base station, which means that the structure can also be used for uplink transmitter-side beam-forming.

An important consequence of the multi-antenna extensions in LTE release 10 was the introduction of an enhanced downlink *reference-signal structure* that more extensively separated the function of channel estimation and the function of acquiring channel-state information. The aim of this was to better enable novel antenna arrangements and new features such as more elaborate multi-point coordination/transmission in a flexible way.

7.3.3 Relaying

In the context of LTE, *relaying* implies that the terminal communicates with the network via a *relay node* that is *wirelessly connected* to a *donor cell* using the LTE radio-interface technology (see Figure 7.5). From a terminal point of view, the relay node will appear as an ordinary cell. This has the important advantage of simplifying the terminal implementation and making the relay node backwards compatible—that is, LTE release-8/9 terminals can access the network via the relay node. In essence, the relay is a low-power base station wirelessly connected to the remaining part of the network.

7.3.4 Heterogeneous deployments

Heterogeneous deployments refer to deployments with a mixture of network nodes with different transmit power and overlapping geographical coverage (see Figure 7.6). A typical

Copyright Elsevier 2024

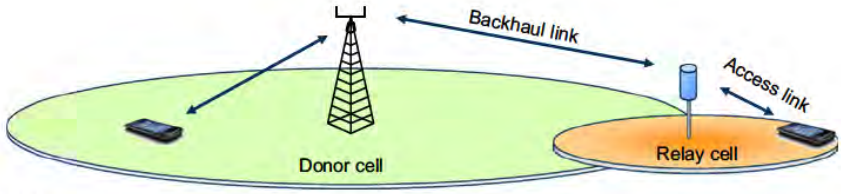


FIGURE 7.5

Example of relaying

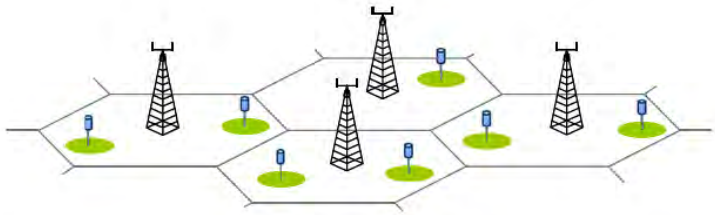


FIGURE 7.6

Example of heterogeneous deployment with low-power nodes inside macro cells

example is a pico node placed within the coverage area of a macro cell. Although such deployments were already supported in release 8, release 10 introduced new means to handle the inter-layer interference that may occur between, for example, a pico layer and the overlaid macro layer. The multi-point-coordination techniques introduced in release 11 further extend the set of tools for supporting heterogeneous deployments.

7.4 Release 11

Release 11 further extended the performance and capabilities of LTE. One of the most notable features of LTE release 11, finalized in late 2012, was radio-interface functionality for multi-point coordination and transmission. Other examples of improvements in release 11 were carrier-aggregation enhancements, a new control-channel structure, and performance requirements for more advanced terminal receivers.

In all new 3GPP releases, there is support added for operation in additional frequency bands. In release 11, there was not only support added for new bands, but also requirements defined for carrier-aggregation within and between both existing and new bands. In line with the evolution of RF technology to support larger transmitter and receiver bandwidths, there were also requirements added in the specifications for

Copyright Elsevier 2024

multiband base stations, supporting simultaneous transmission and/or reception of carriers in multiple frequency bands using the same RF hardware. The RF requirements are described in more detail in Chapter 19.

7.4.1 Multi-point coordination and transmission

As discussed earlier, the first release of LTE included some specific support for coordination between transmission points, referred to as *Inter-Cell Interference Coordination* (ICIC), as a means to at least partly control the interference between cells. However, the support for such coordination was significantly expanded as part of LTE release 11, including the possibility for much more dynamic coordination between transmission points.

In contrast to release-8 ICIC, which was limited to the definition of certain X2 messages to assist coordination between cells, the release-11 activities focused on radio-interface features and terminal functionality to assist different coordination means, including the support for channel-state feedback for multiple transmission points. Jointly these features and functionality go under the name *Coordinated Multi-Point* (CoMP) transmission/reception. Refinement to the reference-signal structure were also an important part of the CoMP support, as was the enhanced control-channel structure introduced as part of release 11 discussed further below.

Support for CoMP includes *multi-point coordination*—that is, when transmission to a terminal is carried out from one specific transmission point, but where scheduling and link adaptation is coordinated between the transmission points, as well as *multi-point transmission* in which case transmission to a terminal can be carried out from multiple transmission points either in such a way that that transmission can switch dynamically between different transmission points (*Dynamic Point Selection*) or be carried out jointly from multiple transmission points (*Joint Transmission*).

A similar distinction can be made for uplink where one can distinguish between (uplink) multi-point coordination and multi-point *reception*. In general, though, uplink CoMP is mainly a network implementation issue and has very little impact on the terminal and very little visibility in the radio-interface specifications.

7.4.2 Enhanced control channel structure

In release 11, a new complementary control channel structure was introduced in order to support intercell interference coordination and to exploit the additional flexibility of the new reference-signal structure not only for data transmission, which was the case in release 10, but also for control signaling. The new control-channel structure can thus be seen as a prerequisite for many CoMP schemes, although it is also beneficial for beam-forming and frequency-domain interference coordination as well.

Copyright Elsevier 2024

7.4.3 Carrier-aggregation enhancements

Carrier aggregation was introduced in release 10, but for TDD it was restricted to the case of all component carriers having the same downlink-uplink allocations. In release 11, this restriction was removed, and aggregation of multiple TDD carriers with different downlink-uplink allocations is supported. The motivation was to improve the coexistence with other radio-access technologies in case of carrier aggregation. An interesting aspect of aggregating different downlink-uplink allocations is that the terminal may need to receive and transmit simultaneously in order to fully utilize both carriers. Thus, unlike previous releases, a TDD-capable terminal may, similarly to a FDD-capable terminal, need a duplex filter.

While RF requirements for carrier aggregation were defined only for a very limited set of frequency bands in release 10, this has been extended to cover a number of bands in release 11. Requirements for intra-band carrier aggregation are defined in several bands, while an even larger number of bands have requirements defined for inter-band carrier aggregation. There are also requirements defined for non-contiguous carrier aggregation, where in particular the requirements to support base stations operating in non-contiguous spectrum allocations have been added in.

7.4.4 Advanced receivers

Fundamentally, a terminal vendor is free to design the terminal receiver in any way as long as it supports the minimum requirements defined in the specifications. There is an incentive for the vendors to provide significantly better receivers, as this could be directly translated into improved end-user data rates. However, the network may not be able to exploit such receiver improvements to their full extent, as it might not know which terminals have significantly better performance. Network deployments therefore need to be based on the minimum requirements. Defining performance requirements for more advanced receiver types to some extent alleviates this, as the minimum performance of a terminal equipped with an advanced receiver is known.

7.5 Terminal capabilities

To support different scenarios, which may call for different terminal capabilities in terms of data rates, as well as to allow for market differentiation in terms of low- and high-end terminals with a corresponding difference in price, not all terminals support all capabilities. Furthermore, terminals from an earlier release of the standard will not support features introduced in later versions of LTE. For example, a release-8 terminal will not support carrier aggregation as this feature was introduced in release 10. Therefore, as part of the connection

Copyright Elsevier 2024

Table 7.1 UE Categories

	Category							
	Release 8/9/10/11					Release 10/11 only		
	1	2	3	4	5	6	7	8
Downlink peak rate [Mbit/s]	10	50	100	150	300	300	300	3000
Uplink peak rate [Mbit/s]	5	25	50	50	75	50	150	1500
Maximum downlink modulation	64QAM							
Maximum uplink modulation	16QAM				64QAM	16QAM	64QAM	
Max number of layers for downlink spatial multiplexing	1	2			4	Signaled separately		

setup, the terminal indicates not only which release of LTE it supports, but also its capabilities within the release.

In principle, the different parameters could be specified separately, but to limit the number of combinations and avoid a parameter combination that does not make sense, a set of physical-layer capabilities are lumped together to form a UE category (UE, *User Equipment*, is the term used in 3GPP to denote a terminal). In total five different UE categories have been specified for LTE release 8/9, ranging from the low-end category 1 not supporting spatial multiplexing to the high-end category 5 supporting the full set of features in the release-8/9 physical layer specifications. The categories are summarized in Table 7.1 (in simplified form; for the full set of details, see [67]). Note that, regardless of the category, a terminal is always capable of receiving single-stream transmissions from up to four antenna ports. This is necessary as the system information can be transmitted on up to four antenna ports.

In LTE release 10, features such as carrier aggregation and uplink spatial multiplexing were introduced, which called for additional capability signaling compared to release 8/9, either in the form of additional UE categories or as separate capabilities. Defining new categories for each foreseen combination of the maximum number of component carriers and maximum degree of spatial multiplexing could be done in principle, although the number of categories might become very large and which categories a terminal supports may be frequency-band dependent. Therefore, in release 10, three additional UE categories were defined as seen in Table 7.1, and the maximum number of component carriers and degree of spatial multiplexing supported, both in uplink and downlink, are signaled separately from the category number. A release-10/11 terminal may therefore declare itself as, for example, category 4 but capable of uplink spatial multiplexing. Hence, categories 1–5 may have a

Copyright Elsevier 2024

slightly different meaning for a release-8/9 and a release-10/11 terminal, depending on the value of the separately declared capabilities. Furthermore, in order to be able to operate in release-8/9 networks, a release-10/11 terminal has to be able to declare both release-8/9 and release-10 categories.

In addition to the capabilities mentioned in the different UE categories, there are some capabilities specified outside the categories. The duplexing schemes supported is one such example, and the support of UE-specific reference signals for FDD in release 8 is another. Whether the terminal supports other radio-access technologies, for example GSM and WCDMA, is also declared separately.

Copyright Elsevier 2024

This page intentionally left blank

Copyright Elsevier 202

CHAPTER

Radio-Interface Architecture

8

This chapter contains a brief overview of the overall architecture of an LTE radio-access network and the associated core network, followed by descriptions of the radio-access network user-plane and control-plane protocols.

8.1 Overall system architecture

In parallel to the work on the LTE radio-access technology in 3GPP, the overall system architecture of both the *Radio-Access Network* (RAN) and the *Core Network* (CN) was revisited, including the split of functionality between the two network parts. This work was known as the *System Architecture Evolution* (SAE) and resulted in a flat RAN architecture, as well as a new core-network architecture referred to as the *Evolved Packet Core* (EPC). Together, the LTE RAN and the EPC are referred to as the *Evolved Packet System* (EPS).[1]

The RAN is responsible for all radio-related functionality of the overall network including, for example, scheduling, radio-resource handling, retransmission protocols, coding and various multi-antenna schemes. These functions will be discussed in detail in the subsequent chapters.

The EPC is responsible for functions not related to the radio access but needed for providing a complete mobile-broadband network. This includes, for example, authentication, charging functionality, and setup of end-to-end connections. Handling these functions separately, instead of integrating them into the RAN, is beneficial as it allows for several radio-access technologies to be served by the same core network.

Although this book focuses on the LTE RAN, a brief overview of the EPC, as well as how it connects to the RAN, is useful. For an excellent in-depth discussion of EPC, the reader is referred to [9].

8.1.1 Core network

The EPC is a radical evolution from the GSM/GPRS core network used for GSM and WCDMA/HSPA. EPC supports access to the *packet-switched domain* only, with no access to

[1]UTRAN, the WCDMA/HSPA radio-access network, is also part of the EPS.

4G: LTE/LTE-Advanced for Mobile Broadband. http://dx.doi.org/10.1016/B978-0-12-419985-9.00008-8
Copyright © 2014 Erik Dahlman, Stefan Parkvall and Johan Sköld. Published by Elsevier Ltd. All rights reserved.

Copyright Elsevier 2024

Samsung Exhibit 1017, Page 147 of 538
Samsung Electronics Co., Ltd. v. Asus Technology Licensing Inc.
IPR2024-01217

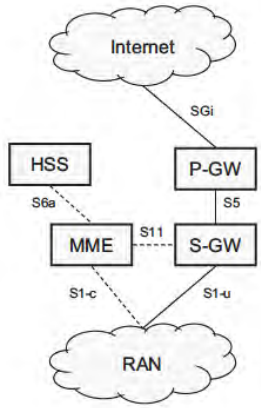


FIGURE 8.1

Core network architecture

the *circuit-switched domain*. It consists of several different types of nodes, some of which are briefly described below and illustrated in Figure 8.1.

The *Mobility Management Entity* (MME) is the control-plane node of the EPC. Its responsibilities include connection/release of bearers to a terminal, handling of IDLE to ACTIVE transitions, and handling of security keys. The functionality operating between the EPC and the terminal is sometimes referred to as the *Non-Access Stratum* (NAS), to separate it from the *Access Stratum* (AS) which handles functionality operating between the terminal and the radio-access network.

The *Serving Gateway* (S-GW) is the user-plane node connecting the EPC to the LTE RAN. The S-GW acts as a mobility anchor when terminals move between eNodeBs (see next section), as well as a mobility anchor for other 3GPP technologies (GSM/GPRS and HSPA). Collection of information and statistics necessary for charging is also handled by the S-GW.

The *Packet Data Network Gateway* (PDN Gateway, P-GW) connects the EPC to the internet. Allocation of the IP address for a specific terminal is handled by the P-GW, as well as quality-of-service enforcement according to the policy controlled by the PCRF (see below). The P-GW is also the mobility anchor for non-3GPP radio-access technologies, such as CDMA2000, connected to the EPC.

In addition, the EPC also contains other types of nodes such as *Policy and Charging Rules Function* (PCRF) responsible for quality-of-service (QoS) handling and charging, and the *Home Subscriber Service* (HSS) node, a database containing subscriber information. There are also some additional nodes present as regards to network support of *Multimedia Broadcast Multicast Services* (MBMS) (see Chapter 17 for a more detailed description of MBMS, including the related architecture aspects).

Copyright Elsevier 2024

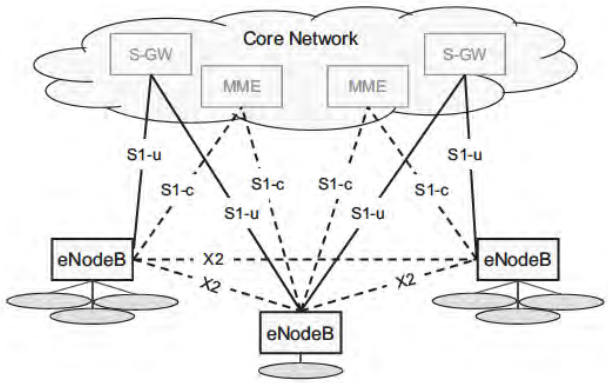


FIGURE 8.2

Radio-access network interfaces

It should be noted that the nodes discussed above are *logical* nodes. In an actual physical implementation, several of them may very well be combined. For example, the MME, P-GW, and S-GW could very well be combined into a single physical node.

8.1.2 Radio-access network

The LTE radio-access network uses a flat architecture with a single type of node[2]—the *eNodeB*. The eNodeB is responsible for all radio-related functions in one or several cells. It is important to note that an eNodeB is a *logical* node and not a physical implementation. One common implementation of an eNodeB is a three-sector site, where a base station is handling transmissions in three cells, although other implementations can be found as well, such as one baseband processing unit to which a number of remote radio heads are connected. One example of the latter is a large number of indoor cells, or several cells along a highway, belonging to the same eNodeB. Thus, a base station is a *possible* implementation of, but not *the same* as, an eNodeB.

As can be seen in Figure 8.2, the eNodeB is connected to the EPC by means of the *S1 interface*, more specifically to the S-GW by means of the *S1 user-plane part*, S1-u, and to the MME by means of the *S1 control-plane part*, S1-c. One eNodeB can be connected to multiple MMEs/S-GWs for the purpose of load sharing and redundancy.

The *X2 interface*, connecting eNodeBs to each other, is mainly used to support active-mode mobility. This interface may also be used for multi-cell *Radio Resource Management* (RRM) functions such as *Inter-Cell Interference Coordination* (ICIC) discussed in

[2]The introduction of MBMS (see Chapter 17) in release 9 and relaying (see Chapter 18) in release 10 brings additional node types to the RAN.

Copyright Elsevier 2024

Chapter 15. The X2 interface is also used to support lossless mobility between neighboring cells by means of packet forwarding.

8.2 Radio protocol architecture

With the overall network architecture in mind, the RAN protocol architecture for the user and control planes can be discussed. Figure 8.3 illustrates the RAN protocol architecture (the MME is, as discussed in the previous section, not part of the RAN but is included in the figure for completeness). As seen in the figure, many of the protocol entities are common to the user and control planes. Therefore, although this section mainly describes the protocol architecture from a user-plane perspective, the description is in many respects also applicable to the control plane. Control-plane-specific aspects are discussed in Section 8.3.

The LTE radio-access network provides one or more *Radio Bearers* to which IP packets are mapped according to their Quality-of-Service requirements. A general overview of the LTE (user-plane) protocol architecture for the downlink is illustrated in Figure 8.4. As will become clear in the subsequent discussion, not all the entities illustrated in Figure 8.4 are applicable in all situations. For example, neither MAC scheduling nor hybrid ARQ with soft combining is used for broadcast of the basic system information. The LTE protocol structure related to uplink transmissions is similar to the downlink structure in Figure 8.4, although there are some differences with respect to, for example, transport-format selection.

The different protocol entities of the radio-access network are summarized below and described in more detail in the following sections.

- *Packet Data Convergence Protocol* (PDCP) performs IP header compression to reduce the number of bits to transmit over the radio interface. The header-compression mechanism is based on Robust Header Compression (ROHC) [68], a standardized

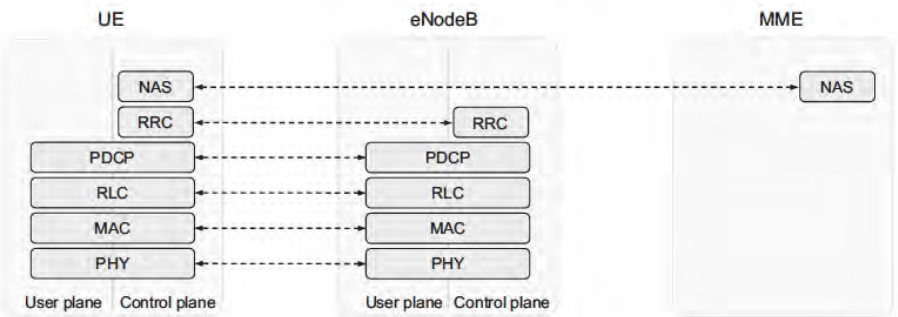


FIGURE 8.3

Overall RAN protocol architecture

Copyright Elsevier 2024

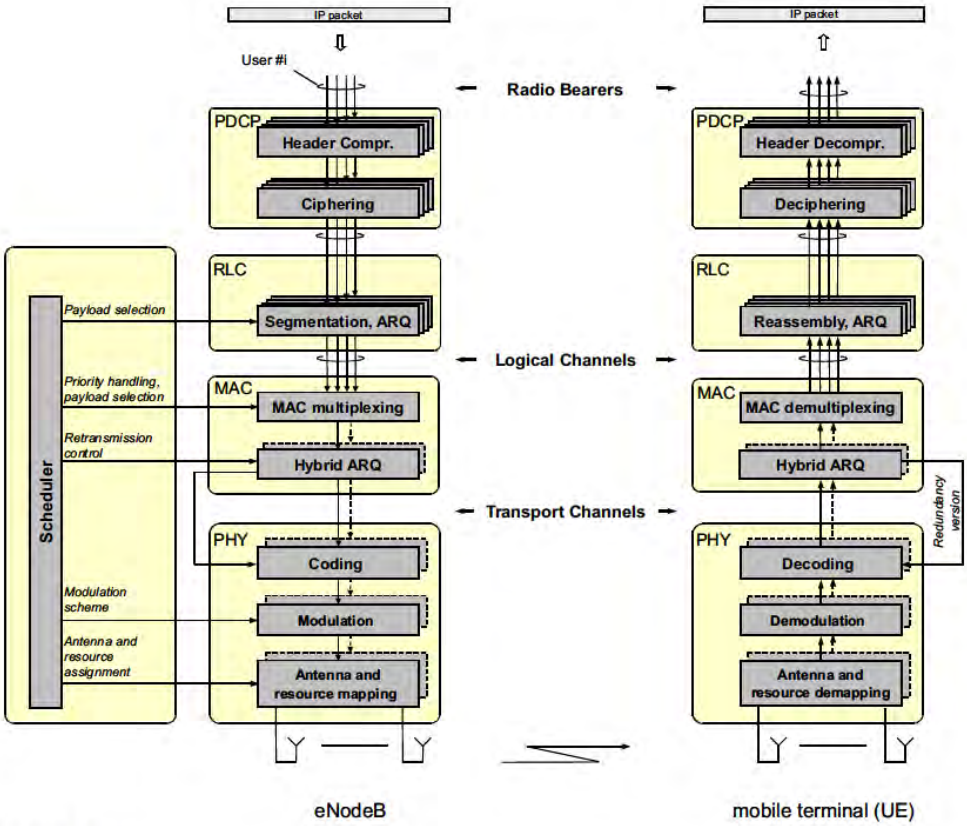


FIGURE 8.4

LTE protocol architecture (downlink)

header-compression algorithm also used for several other mobile-communication technologies. PDCP is also responsible for ciphering and, for the control plane, integrity protection of the transmitted data, as well as in-sequence delivery and duplicate removal for handover. At the receiver side, the PDCP protocol performs the corresponding deciphering and decompression operations. There is one PDCP entity per radio bearer configured for a terminal.

- *Radio-Link Control* (RLC) is responsible for segmentation/concatenation, retransmission handling, duplicate detection, and in-sequence delivery to higher layers. The RLC provides services to the PDCP. There is one RLC entity per radio bearer configured for a terminal.

Copyright Elsevier 2024

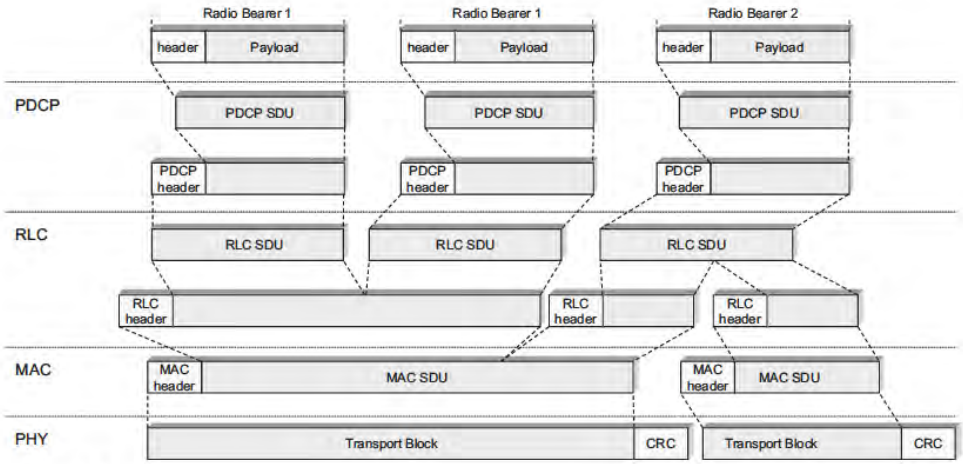


FIGURE 8.5

Example of LTE data flow

- *Medium-Access Control* (MAC) handles multiplexing of logical channels, hybrid-ARQ retransmissions, and uplink and downlink scheduling. The scheduling functionality is located in the eNodeB for both uplink and downlink. The hybrid-ARQ protocol part is present in both the transmitting and receiving ends of the MAC protocol. The MAC provides services to the RLC in the form of *logical channels*.
- *Physical Layer* (PHY) handles coding/decoding, modulation/demodulation, multi-antenna mapping, and other typical physical-layer functions. The physical layer offers services to the MAC layer in the form of *transport channels*.

To summarize the flow of downlink data through all the protocol layers, an example illustration for a case with three IP packets, two on one radio bearer and one on another radio bearer, is given in Figure 8.5. The data flow in the case of uplink transmission is similar. The PDCP performs (optional) IP-header compression, followed by ciphering. A PDCP header is added, carrying information required for deciphering in the terminal. The output from the PDCP is forwarded to the RLC.

The RLC protocol performs concatenation and/or segmentation of the PDCP SDUs[3] and adds an RLC header. The header is used for in-sequence delivery (per logical channel) in the terminal and for identification of RLC PDUs in the case of retransmissions. The RLC PDUs are forwarded to the MAC layer, which multiplexes a number of RLC PDUs and attaches

[3]In general, the data entity from/to a higher protocol layer is known as a Service Data Unit (SDU), and the corresponding entity to/from a lower protocol layer entity is called a Protocol Data Unit (PDU).

Copyright Elsevier 2024

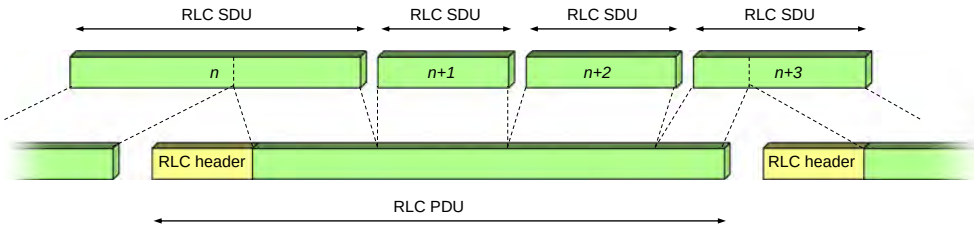


FIGURE 8.6

RLC segmentation and concatenation

a MAC header to form a transport block. The transport-block size depends on the instantaneous data rate selected by the link-adaptation mechanism. Thus, the link adaptation affects both the MAC and RLC processing. Finally, the physical layer attaches a CRC to the transport block for error-detection purposes, performs coding and modulation, and transmits the resulting signal, possibly using multiple transmit antennas.

The remainder of the chapter contains an overview of the RLC, MAC, and physical layers. A more detailed description of the LTE physical-layer processing is given in Chapters 10 (downlink) and 11 (uplink), followed by descriptions of some specific uplink and downlink radio-interface functions and procedures in the subsequent chapters.

8.2.1 Radio-link control

The RLC protocol is responsible for segmentation/concatenation of (header-compressed) IP packets, also known as RLC SDUs, from the PDCP into suitably sized RLC PDUs. It also handles retransmission of erroneously received PDUs, as well as removal of duplicated PDUs. Finally, the RLC ensures in-sequence delivery of SDUs to upper layers. Depending on the type of service, the RLC can be configured in different modes to perform some or all of these functions.

Segmentation and concatenation, one of the main RLC functions, is illustrated in Figure 8.6. Depending on the scheduler decision, a certain amount of data is selected for transmission from the RLC SDU buffer and the SDUs are segmented/concatenated to create the RLC PDU. Thus, for LTE the RLC PDU size varies *dynamically*. For high data rates, a large PDU size results in a smaller relative overhead, while for low data rates, a small PDU size is required as the payload would otherwise be too large. Hence, as the LTE data rates may range from a few kbit/s up to 3 Gbit/s, dynamic PDU sizes are motivated for LTE in contrast to earlier mobile-communication technologies, which typically used a fixed PDU size. Since the RLC, scheduler, and rate-adaptation mechanisms are all located in the eNodeB, dynamic PDU sizes are easily supported for LTE. In each RLC PDU, a header is included, containing, among other things, a sequence number used for in-sequence delivery and by the retransmission mechanism.

Copyright Elsevier 2024

The RLC retransmission mechanism is also responsible for providing error-free delivery of data to higher layers. To accomplish this, a retransmission protocol operates between the RLC entities in the receiver and transmitter. By monitoring the sequence numbers of the incoming PDUs, the receiving RLC can identify missing PDUs. Status reports are then fed back to the transmitting RLC entity, requesting retransmission of missing PDUs. Based on the received status report, the RLC entity at the transmitter can take the appropriate action and retransmit the missing PDUs if needed.

Although the RLC is capable of handling transmission errors due to noise, unpredictable channel variations, etc., error-free delivery is in most cases handled by the MAC-based hybrid-ARQ protocol. The use of a retransmission mechanism in the RLC may therefore seem superfluous at first. However, as will be discussed in Section 8.2.2.3, this is not the case and the use of both RLC- and MAC-based retransmission mechanisms is in fact well motivated by the differences in the feedback signaling.

The details of RLC are further described in Chapter 12.

8.2.2 Medium-access control

The MAC layer handles logical-channel multiplexing, hybrid-ARQ retransmissions, and uplink and downlink scheduling. It is also responsible for multiplexing/demultiplexing data across multiple component carriers when carrier aggregation is used.

8.2.2.1 Logical channels and transport channels

The MAC provides services to the RLC in the form of *logical channels*. A logical channel is defined by the *type* of information it carries and is generally classified as a *control channel*, used for transmission of control and configuration information necessary for operating an LTE system, or as a *traffic channel*, used for the user data. The set of logical-channel types specified for LTE includes:

- *The Broadcast Control Channel* (BCCH), used for transmission of *system information* from the network to all terminals in a cell. Prior to accessing the system, a terminal needs to acquire the system information to find out how the system is configured and, in general, how to behave properly within a cell.
- The *Paging Control Channel* (PCCH), used for paging of terminals whose location on a cell level is not known to the network. The paging message therefore needs to be transmitted in multiple cells.
- The *Common Control Channel* (CCCH), used for transmission of control information in conjunction with random access.
- The *Dedicated Control Channel* (DCCH), used for transmission of control information to/from a terminal. This channel is used for individual configuration of terminals such as different handover messages.
- The *Multicast Control Channel* (MCCH), used for transmission of control information required for reception of the MTCH (see below).

Copyright Elsevier 2024

- The *Dedicated Traffic Channel* (DTCH), used for transmission of user data to/from a terminal. This is the logical channel type used for transmission of all uplink and non-MBSFN downlink user data.
- The *Multicast Traffic Channel* (MTCH), used for downlink transmission of MBMS services.

From the physical layer, the MAC layer uses services in the form of *transport channels*. A transport channel is defined by *how* and *with what characteristics* the information is transmitted over the radio interface. Data on a transport channel is organized into *transport blocks*. In each *Transmission Time Interval* (TTI), at most one transport block of dynamic size is transmitted over the radio interface to/from a terminal in the absence of spatial multiplexing. In the case of spatial multiplexing (MIMO), there can be up to two transport blocks per TTI.

Associated with each transport block is a *Transport Format* (TF), specifying *how* the transport block is to be transmitted over the radio interface. The transport format includes information about the transport-block size, the modulation-and-coding scheme, and the antenna mapping. By varying the transport format, the MAC layer can thus realize different data rates. Rate control is also known as *transport-format selection*.

The following transport-channel types are defined for LTE:

- The *Broadcast Channel* (BCH) has a fixed transport format, provided by the specifications. It is used for transmission of parts of the BCCH system information, more specifically the so-called *Master Information Block* (MIB), as described in Chapter 14.
- The *Paging Channel* (PCH) is used for transmission of paging information from the PCCH logical channel. The PCH supports *discontinuous reception* (DRX) to allow the terminal to save battery power by waking up to receive the PCH only at predefined time instants. The LTE paging mechanism is also described in Chapter 14.
- The *Downlink Shared Channel* (DL-SCH) is the main transport channel used for transmission of downlink data in LTE. It supports key LTE features such as dynamic rate adaptation and channel-dependent scheduling in the time and frequency domains, hybrid ARQ with soft combining, and spatial multiplexing. It also supports DRX to reduce terminal power consumption while still providing an always-on experience. The DL-SCH is also used for transmission of the parts of the BCCH system information not mapped to the BCH. There can be multiple DL-SCHs in a cell, one per terminal[4] scheduled in this TTI, and, in some subframes, one DL-SCH carrying system information.
- The *Multicast Channel* (MCH) is used to support MBMS. It is characterized by a semi-static transport format and semi-static scheduling. In the case of multi-cell transmission using MBSFN, the scheduling and transport format configuration is

[4]For carrier aggregation, a terminal may receive multiple DL-SCHs, one per component carrier.

Copyright Elsevier 2024

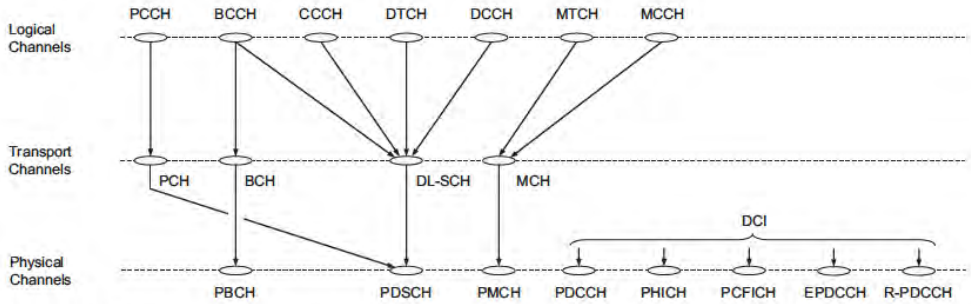


FIGURE 8.7

Downlink channel mapping

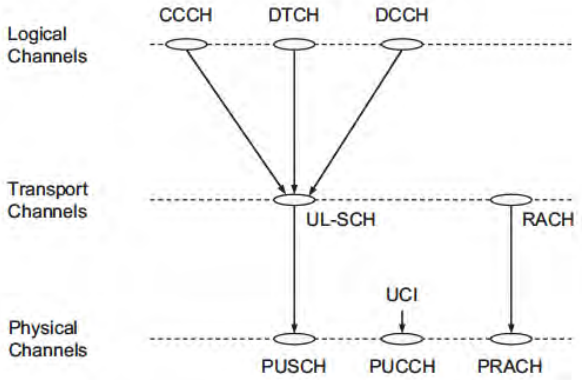


FIGURE 8.8

Uplink channel mapping

coordinated among the transmission points involved in the MBSFN transmission. MBSFN transmission is described in Chapter 17.

- The Uplink Shared Channel (UL-SCH) is the uplink counterpart to the DL-SCH—that is, the uplink transport channel used for transmission of uplink data.

In addition, the *Random-Access Channel* (RACH) is also defined as a transport channel, although it does not carry transport blocks.

Part of the MAC functionality is multiplexing of different logical channels and mapping of the logical channels to the appropriate transport channels. The mapping between logical-channel types and transport channel types is given in Figure 8.7 for the downlink and Figure 8.8 for the uplink. The figures clearly indicate how DL-SCH and UL-SCH are the main

Copyright Elsevier 2024

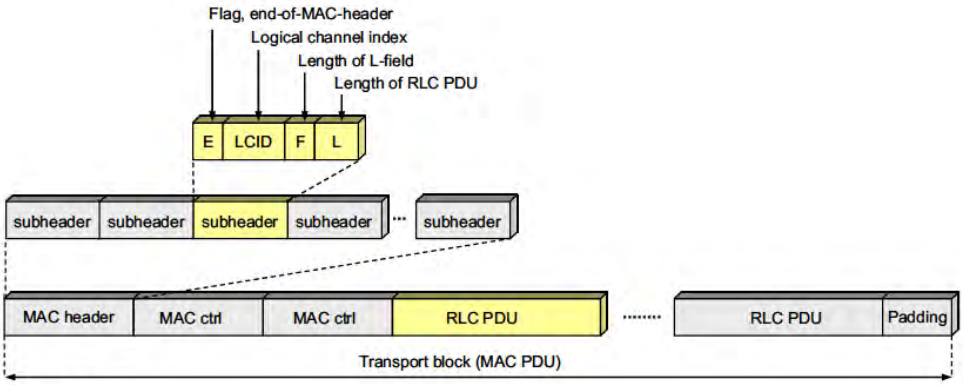


FIGURE 8.9

MAC header and SDU multiplexing

downlink and uplink transport channels respectively. In the figures, the corresponding physical channels, described further below, are also included and the mapping between transport channels and physical channels is illustrated.

To support priority handling, multiple logical channels, where each logical channel has its own RLC entity, can be multiplexed into one transport channel by the MAC layer. At the receiver, the MAC layer handles the corresponding demultiplexing and forwards the RLC PDUs to their respective RLC entity for in-sequence delivery and the other functions handled by the RLC. To support the demultiplexing at the receiver, a MAC header, shown in Figure 8.9, is used. To each RLC PDU, there is an associated sub-header in the MAC header. The sub-header contains the identity of the logical channel (LCID) from which the RLC PDU originated and the length of the PDU in bytes. There is also a flag indicating whether this is the last sub-header or not. One or several RLC PDUs, together with the MAC header and, if necessary, padding to meet the scheduled transport-block size, form one transport block which is forwarded to the physical layer.

In addition to multiplexing of different logical channels, the MAC layer can also insert the so-called *MAC control elements* into the transport blocks to be transmitted over the transport channels. A MAC control element is used for inband control signaling—for example, timing-advance commands and random-access response, as described in Sections 11.6 and 14.3 respectively. Control elements are identified with reserved values in the LCID field, where the LCID value indicates the type of control information. Furthermore, the length field in the sub-header is removed for control elements with a fixed length.

The MAC multiplexing functionality is also responsible for handling of multiple component carriers in the case of carrier aggregation. The basic principle for carrier aggregation is independent processing of the component carriers in the physical layer, including

Copyright Elsevier 2024

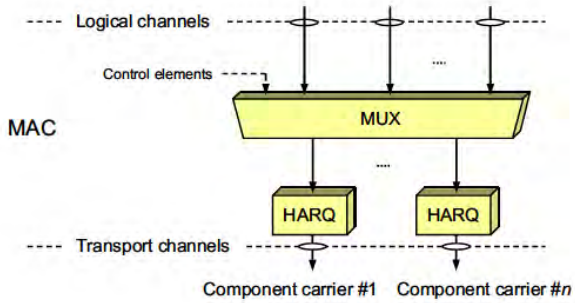


FIGURE 8.10

MAC operation in case of multiple component carriers

control signaling, scheduling, and hybrid-ARQ retransmissions, while carrier aggregation is invisible to RLC and PDCP. Carrier aggregation is therefore mainly seen in the MAC layer, as illustrated in Figure 8.10, where logical channels, including any MAC control elements, are multiplexed to form one (two in the case of spatial multiplexing) transport block(s) per component carrier with each component carrier having its own hybrid-ARQ entity.

8.2.2.2 *Scheduling*

One of the basic principles of LTE radio access is shared-channel transmission—that is, time–frequency resources are dynamically shared between users. The *scheduler* is part of the MAC layer (although often better viewed as a separate entity, as illustrated in Figure 8.4) and controls the assignment of uplink and downlink resources in terms of so-called *resource-block pairs*. Resource-block pairs correspond to a time–frequency unit of 1 ms times 180 kHz, as described in more detail in Chapter 9.

The basic operation of the scheduler is so-called *dynamic* scheduling, where the eNodeB in each 1 ms interval takes a scheduling decision and sends scheduling information to the selected set of terminals. However, there is also a possibility for semi-persistent scheduling where a semi-static scheduling pattern is signaled in advance to reduce the control-signaling overhead.

Uplink and downlink scheduling are separated in LTE, and uplink and downlink scheduling decisions can be taken independently of each other (within the limits set by the uplink/downlink split in the case of half-duplex FDD operation).

The downlink scheduler is responsible for (dynamically) controlling which terminal(s) to transmit to and, for each of these terminals, the set of resource blocks upon which the terminal's DL-SCH should be transmitted. Transport-format selection (selection of transport-block size, modulation scheme, and antenna mapping) and logical-channel multiplexing for downlink transmissions are controlled by the eNodeB, as illustrated in the left part of Figure 8.11. As a consequence of the scheduler controlling the data rate, the RLC segmentation and MAC multiplexing will also be affected by the scheduling decision. The outputs from the downlink scheduler can be seen in Figure 8.4.

Copyright Elsevier 2024

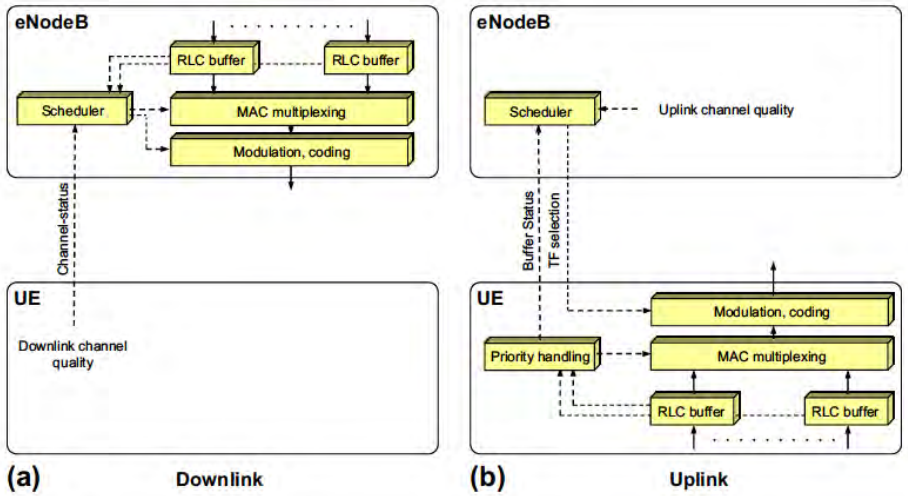


FIGURE 8.11

Transport-format selection in (a) downlink and (b) uplink

The uplink scheduler serves a similar purpose, namely to (dynamically) control which terminals are to transmit on their respective UL-SCH and on which uplink time–frequency resources (including component carrier). Despite the fact that the eNodeB scheduler determines the transport format for the terminal, it is important to point out that the uplink scheduling decision is taken *per terminal* and not per radio bearer. Thus, although the eNodeB scheduler controls the payload of a scheduled terminal, the terminal is still responsible for selecting *from which radio bearer(s)* the data is taken. The terminal autonomously handles logical-channel multiplexing according to predefined rules, the parameters of which can be configured by the eNodeB. This is illustrated in the right part of Figure 8.11, where the eNodeB scheduler controls the transport format and the terminal controls the logical-channel multiplexing.

Although the scheduling strategy is implementation specific and not specified by 3GPP, the overall goal of most schedulers is to take advantage of the channel variations between terminals and preferably schedule transmissions to a terminal on resources with advantageous channel conditions. A benefit of the use of OFDM in LTE is the possibility to exploit channel variations in both time and frequency domains through channel-dependent scheduling. This was mentioned earlier in Chapter 7 and illustrated in Figure 7.2. For the larger bandwidths supported by LTE, where a significant amount of frequency-selective fading may occur, the possibility for the scheduler to also exploit frequency-domain channel variations becomes increasingly important compared to exploiting time-domain variations only. This is beneficial

Copyright Elsevier 2024

especially at low speeds, where the variations in the time domain are relatively slow compared to the delay requirements set by many services.

Downlink channel-dependent scheduling is supported through *channel-state information* (CSI), reported by the terminal to the eNodeB and reflecting the instantaneous downlink channel quality in the time and frequency domains, as well as information necessary to determine the appropriate antenna processing in the case of spatial multiplexing. In the uplink, the channel-state information necessary for uplink channel-dependent scheduling can be based on a *sounding reference signal* transmitted from each terminal for which the eNodeB wants to estimate the uplink channel quality. To aid the uplink scheduler in its decisions, the terminal can transmit buffer-status information to the eNodeB using a MAC message. This information can only be transmitted if the terminal has been given a valid scheduling grant. For situations when this is not the case, an indicator that the terminal needs uplink resources is provided as part of the uplink L1/L2 control-signaling structure (see Chapter 11).

8.2.2.3 Hybrid ARQ with soft combining

Hybrid ARQ with soft combining provides robustness against transmission errors. As hybrid-ARQ retransmissions are fast, many services allow for one or multiple retransmissions, thereby forming an implicit (closed loop) rate-control mechanism. The hybrid-ARQ protocol is part of the MAC layer, while the actual soft combining is handled by the physical layer.[5]

Hybrid ARQ is not applicable for all types of traffic. For example, broadcast transmissions, where the same information is intended for multiple terminals, typically do not rely on hybrid ARQ. Hence, hybrid ARQ is only supported for the DL-SCH and the UL-SCH, although its usage is optional.

The LTE hybrid-ARQ protocol uses multiple parallel stop-and-wait processes. Upon reception of a transport block, the receiver makes an attempt to decode the transport block and informs the transmitter about the outcome of the decoding operation through a single acknowledgement bit, indicating whether the decoding was successful or if a retransmission of the transport block is required. Further details on transmission of hybrid-ARQ acknowledgements are found in Chapters 10—12. The receiver must know to which hybrid-ARQ process a received acknowledgement is associated. This is solved by using the timing of the acknowledgement for association with a certain hybrid-ARQ process. Note that, in the case of TDD operation, the time relation between the reception of data in a certain hybrid-ARQ process and the transmission of the acknowledgement is also affected by the uplink/downlink allocation.

The use of multiple parallel hybrid-ARQ processes, illustrated in Figure 8.12, for each user can result in data being delivered from the hybrid-ARQ mechanism out of sequence. For example, transport block 5 in the figure was successfully decoded before transport block 1,

[5]The soft combining is done before or as part of the channel decoding, which is a physical-layer functionality.

Copyright Elsevier 2024

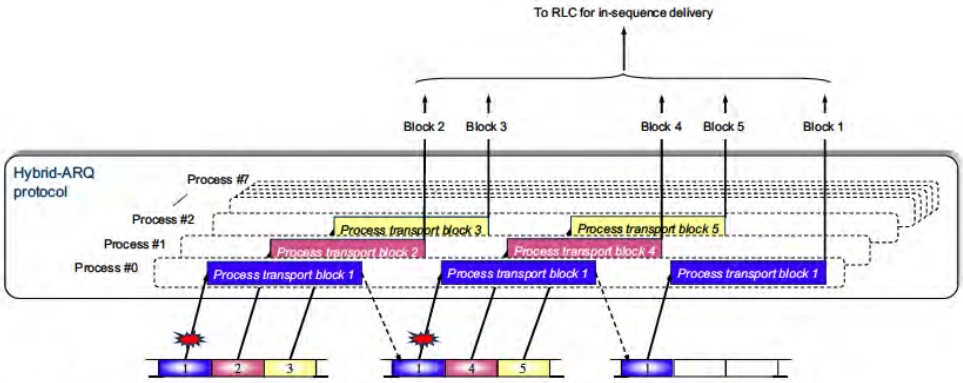


FIGURE 8.12

Multiple parallel hybrid-ARQ processes

which required retransmissions. In-sequence delivery of data is therefore ensured by the RLC layer. The same reordering mechanism in the RLC is also used to handle in-sequence delivery across multiple component carriers in the case of carrier aggregation. As mentioned above, the hybrid-ARQ retransmissions are handled independently per component carrier, which may result in out-of-sequence delivery in a similar way as within a component carrier.

Downlink retransmissions may occur at any time after the initial transmission—that is, the protocol is asynchronous—and an explicit hybrid-ARQ process number is used to indicate which process is being addressed. In an asynchronous hybrid-ARQ protocol, the retransmissions are in principle scheduled similarly to the initial transmissions. Uplink retransmissions, on the other hand, are based on a synchronous protocol, the retransmission occurs at a predefined time after the initial transmission, and the process number can be implicitly derived. In a synchronous protocol the time instant for the retransmissions is fixed once the initial transmission has been scheduled, which must be accounted for in the scheduling operation. However, note that the scheduler knows from the hybrid-ARQ entity in the eNodeB whether a terminal will perform a retransmission or not.

The hybrid-ARQ mechanism will rapidly correct transmission errors due to noise or unpredictable channel variations. As discussed above, the RLC is also capable of requesting retransmissions, which at first sight may seem unnecessary. However, the reason for having two retransmission mechanisms on top of each other can be seen in the feedback signaling. Hybrid ARQ provides fast retransmissions, but due to errors in the feedback, the residual error rate is typically too high for, for example, good TCP performance, while RLC ensures (almost) error-free data delivery but slower retransmissions than the hybrid-ARQ protocol. Hence, the combination of hybrid ARQ and RLC provides an attractive combination of small round-trip time and reliable data delivery. Furthermore, as the RLC and hybrid ARQ are

Copyright Elsevier 2024

located in the same node, tight interaction between the two is possible, as discussed in Chapter 12.

8.2.3 Physical layer

The physical layer is responsible for coding, physical-layer hybrid-ARQ processing, modulation, multi-antenna processing, and mapping of the signal to the appropriate physical time–frequency resources. It also handles mapping of transport channels to physical channels, as shown in Figures 8.7 and 8.8.

As mentioned in the introduction, the physical layer provides services to the MAC layer in the form of transport channels. Data transmission in downlink and uplink use the DL-SCH and UL-SCH transport-channel types respectively. There is at most one or, in the case of spatial multiplexing, two transport blocks per TTI on a DL-SCH or UL-SCH. In the case of carrier aggregation, there is one DL-SCH (or UL-SCH) per component carrier as seen by the terminal.

A *physical channel* corresponds to the set of time–frequency resources used for transmission of a particular transport channel and each transport channel is mapped to a corresponding physical channel, as shown in Figures 8.7 and 8.8. In addition to the physical channels with a corresponding transport channel, there are also physical channels without a corresponding transport channel. These channels, known as L1/L2 control channels, are used for *downlink control information* (DCI), providing the terminal with the necessary information for proper reception and decoding of the downlink data transmission, and *uplink control information* (UCI) used for providing the scheduler and the hybrid-ARQ protocol with information about the situation at the terminal.

The following physical-channel types are defined for LTE:

- The *Physical Downlink Shared Channel* (PDSCH) is the main physical channel used for unicast data transmission, but also for transmission of paging information.
- The *Physical Broadcast Channel* (PBCH) carries part of the system information, required by the terminal in order to access the network.
- The *Physical Multicast Channel* (PMCH) is used for MBSFN transmission.
- The *Physical Downlink Control Channel* (PDCCH) is used for downlink control information, mainly scheduling decisions, required for reception of PDSCH, and for scheduling grants enabling transmission on the PUSCH.
- The *Enhanced Physical Downlink Control Channel* (EPDCCH) was introduced in release 11. It essentially serves the same purpose as the PDCCH, but allows for transmission of the control information in a more flexible way.
- The *Relay Physical Downlink Control Channel* (R-PDCCH) was introduced in release 10 and is used to carry L1/L2 control signaling on the donor-eNodeB-to-relay link.
- The *Physical Hybrid-ARQ Indicator Channel* (PHICH) carries the hybrid-ARQ acknowledgement to indicate to the terminal whether a transport block should be retransmitted or not.

Copyright Elsevier 2024

- The *Physical Control Format Indicator Channel* (PCFICH) is a channel providing the terminals with information necessary to decode the set of PDCCHs. There is only one PCFICH per component carrier.
- The *Physical Uplink Shared Channel* (PUSCH) is the uplink counterpart to the PDSCH. There is at most one PUSCH per uplink component carrier per terminal.
- The *Physical Uplink Control Channel* (PUCCH) is used by the terminal to send hybrid-ARQ acknowledgements, indicating to the eNodeB whether the downlink transport block(s) was successfully received or not, to send channel-state reports aiding downlink channel-dependent scheduling, and for requesting resources to transmit uplink data upon. There is at most one PUCCH per terminal.
- The *Physical Random-Access Channel* (PRACH) is used for random access, as described in Chapter 14.

Note that some of the physical channels, more specifically the channels used for downlink control information (PCFICH, PDCCH, PHICH, EPDCCH, R-PDCCH) and uplink control information (PUCCH), do not have a corresponding transport channel mapped to them.

The remaining downlink transport channels are based on the same general physical-layer processing as the DL-SCH, although with some restrictions in the set of features used. This is especially true for PCH and MCH transport channels. For the broadcast of system information on the BCH, a terminal must be able to receive this information channel as one of the first steps prior to accessing the system. Consequently, the transmission format must be known to the terminals a priori and there is no dynamic control of any of the transmission parameters from the MAC layer in this case. The BCH is also mapped to the physical resource (the OFDM time–frequency grid) in a different way, as described in more detail in Chapter 14.

For transmission of paging messages on the PCH, dynamic adaptation of the transmission parameters can, to some extent, be used. In general, the processing in this case is similar to the generic DL-SCH processing. The MAC can control modulation, the amount of resources, and the antenna mapping. However, as an uplink has not yet been established when a terminal is paged, hybrid ARQ cannot be used as there is no possibility for the terminal to transmit a hybrid-ARQ acknowledgement.

As described in Chapter 3, the MCH is used for MBMS transmissions, typically with single-frequency network operation, by transmitting from multiple cells on the same resources with the same format at the same time. Hence, the scheduling of MCH transmissions must be coordinated between the cells involved and dynamic selection of transmission parameters by the MAC is not possible.

8.3 Control-plane protocols

The control-plane protocols are, among other things, responsible for connection setup, mobility, and security. Control messages transmitted from the network to the terminals can

Copyright Elsevier 2024

originate either from the MME, located in the core network, or from the *Radio Resource Control* (RRC), located in the eNodeB.

NAS control-plane functionality, handled by the MME, includes EPS bearer management, authentication, security, and different idle-mode procedures such as paging. It is also responsible for assigning an IP address to a terminal. For a detailed discussion about the NAS control-plane functionality, see [9].

The RRC is located in the eNodeB and is responsible for handling the RAN-related procedures, including:

- Broadcast of system information necessary for the terminal to be able to communicate with a cell. Acquisition of system information is described in Chapter 14.
- Transmission of paging messages originating from the MME to notify the terminal about incoming connection requests. Paging, discussed further in Chapter 14, is used in the RRC_IDLE state (described further below) when the terminal is not connected to a particular cell. Indication of system-information updates is another use of the paging mechanism, as is public warning systems.
- Connection management, including setting up bearers and mobility within LTE. This includes establishing an RRC context—that is, configuring the parameters necessary for communication between the terminal and the radio-access network.
- Mobility functions such as cell (re)selection.
- Measurement configuration and reporting.
- Handling of UE capabilities; when connection is established the terminal will announce its capabilities as all terminals are not capable of supporting all the functionality described in the LTE specifications, as briefly discussed in Chapter 7.

RRC messages are transmitted to the terminal using *signaling radio bearers* (SRBs), using the same set of protocol layers (PDCP, RLC, MAC and PHY) as described in Section 8.2. The SRB is mapped to the common control channel (CCCH) during establishment of connection and, once a connection is established, to the dedicated control channel (DCCH). Control-plane and user-plane data can be multiplexed in the MAC layer and transmitted to the terminal in the same TTI. The aforementioned MAC control elements can also be used for control of radio resources in some specific cases where low latency is more important than ciphering, integrity protection, and reliable transfer.

8.3.1 State machine

In LTE, a terminal can be in two different states (as illustrated in Figure 8.13), RRC_CONNECTED and RRC_IDLE.

In RRC_CONNECTED, there is an RRC context established—that is, the parameters necessary for communication between the terminal and the radio-access network are known to both entities. The cell to which the terminal belongs is known and an identity of the terminal, the *Cell Radio-Network Temporary Identifier* (C-RNTI), used for signaling purposes

Copyright Elsevier 2024

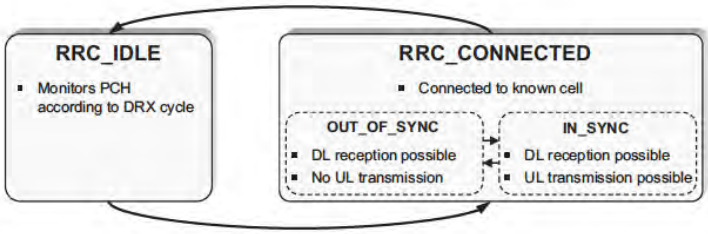


FIGURE 8.13

LTE states

between the terminal and the network, has been configured. RRC_CONNECTED is intended for data transfer to/from the terminal, but *discontinuous reception* (DRX) can be configured in order to reduce terminal power consumption (DRX is described in further detail in Chapter 13). Since there is an RRC context established in the eNodeB in RRC_CONNECTED, leaving DRX and starting to receive/transmit data is relatively fast as no connection with its associated signaling is needed.

Although expressed differently in the specifications, RRC_CONNECTED can be thought of as having two substates, IN_SYNC and OUT_OF_SYNC, depending on whether the uplink is synchronized to the network or not. Since LTE uses an orthogonal FDMA/TDMA-based uplink, it is necessary to synchronize the uplink transmission from different terminals such that they arrive at the receiver at (approximately) the same time. The procedure for obtaining and maintaining uplink synchronization is described in Chapter 11, but in short the receiver measures the arrival time of the transmissions from each actively transmitting terminal and sends timing-correction commands in the downlink. As long as the uplink is synchronized, uplink transmission of user data and L1/L2 control signaling is possible. If no uplink transmission has taken place within a configurable time window, timing alignment is not possible and the uplink is declared to be non-synchronized. In this case, the terminal needs to perform a random-access procedure to restore uplink synchronization prior to transmission of uplink data or control information.

In RRC_IDLE, there is no RRC context in the radio-access network and the terminal does not belong to a specific cell. No data transfer may take place as the terminal sleeps most of the time in order to reduce battery consumption. Uplink synchronization is not maintained and hence the only uplink transmission activity that may take place is random access, discussed in Chapter 14, to move to RRC_CONNECTED. When moving to RRC_CONNECTED the RRC context needs to be established in both the radio-access network and the terminal. Compared to leaving DRX this takes a somewhat longer time. In the downlink, terminals in RRC_IDLE periodically wake up in order to receive paging messages, if any, from the network, as described in Chapter 14.

Copyright Elsevier 2024

This page intentionally left blank

Copyright Elsevier 202

CHAPTER

Physical Transmission Resources

9

In Chapter 8, the overall LTE architecture was discussed, including an overview of the different protocol layers. Prior to discussing the detailed LTE downlink and uplink transmission schemes, a description of the basic time-frequency transmission resource of LTE will be provided in this chapter. An overview of the concept of *antenna ports* will also be provided.

9.1 Overall time–frequency structure

OFDM is the basic transmission scheme for both the downlink and uplink transmission directions in LTE, although, for the uplink, specific means are taken to reduce the cubic metric of the transmitted signal, thereby allowing for improved efficiency for the terminal transmitter power amplifier. Thus, for uplink user data and higher-layer control signaling corresponding to transmission of the PUSCH physical channel, DFT precoding is applied before OFDM modulation, leading to *DFT-spread OFDM* or *DFTS-OFDM*, as described in Chapter 4. As will be described in Chapter 11, for other uplink transmissions, such as the transmission of *L1/L2 control signaling* and different types of *reference-signal transmissions*, other means are taken to limit the cubic metric of the transmitted signal.

The LTE OFDM subcarrier spacing equals 15 kHz for both downlink and uplink. As discussed in Chapter 3, the selection of the subcarrier spacing in an OFDM-based system needs to carefully balance overhead from the cyclic prefix against sensitivity to Doppler spread/shift and other types of frequency errors and inaccuracies. The choice of 15 kHz for the LTE subcarrier spacing was found to offer a good balance between these different constraints.

Assuming an FFT-based transmitter/receiver implementation, 15 kHz subcarrier spacing corresponds to a sampling rate $f_s = 15000\, N_{FFT}$, where N_{FFT} is the FFT size. It is important to understand though that the LTE specifications do not in any way mandate the use of FFT-based transmitter/receiver implementations and even less so a particular FFT size or sampling rate. Nevertheless, FFT-based implementations of OFDM are common practice and an FFT size of 2048, with a corresponding sampling rate of 30.72 MHz, is suitable for the wider LTE carrier bandwidths (such as bandwidths of the order of 15 MHz and above). However, for smaller carrier bandwidths, a smaller FFT size and a correspondingly lower sampling rate can very well be used.

4G: LTE/LTE-Advanced for Mobile Broadband. http://dx.doi.org/10.1016/B978-0-12-419985-9.00009-X
Copyright © 2014 Erik Dahlman, Stefan Parkvall and Johan Sköld. Published by Elsevier Ltd. All rights reserved.

Copyright Elsevier 2024

Samsung Exhibit 1017, Page 167 of 538
Samsung Electronics Co., Ltd. v. Asus Technology Licensing Inc.
IPR2024-01217

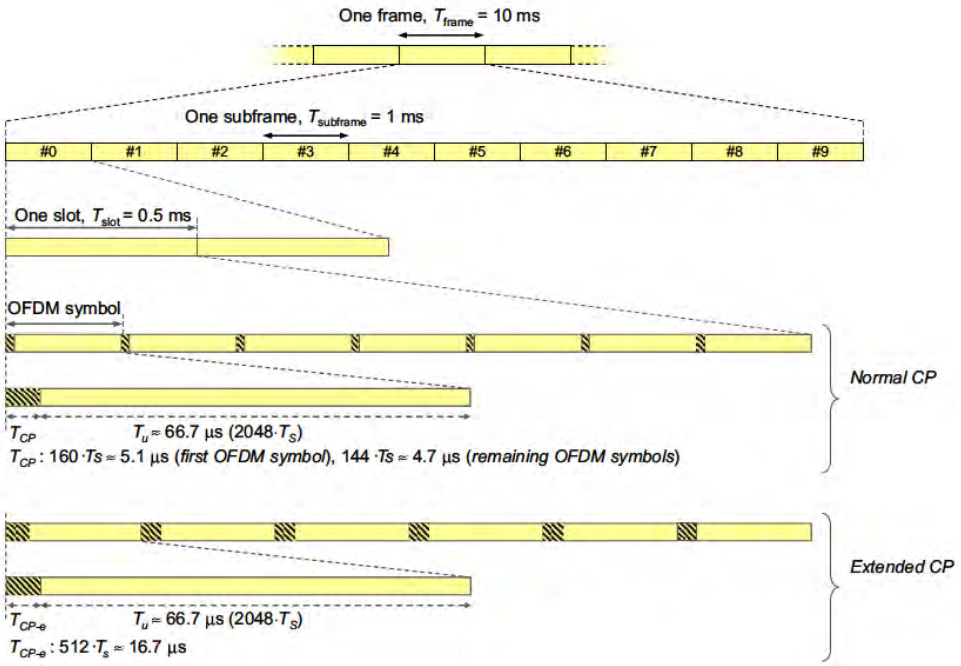


FIGURE 9.1

LTE time-domain structure

In addition to the 15 kHz subcarrier spacing, a *reduced subcarrier spacing* of 7.5 kHz, with a corresponding OFDM symbol time that is twice as long, is also defined for LTE. The introduction of the reduced subcarrier spacing specifically targeted MBSFN-based multicast/broadcast transmissions (see Chapter 17). However, currently the 7.5 kHz subcarrier numerology is only partly implemented in the LTE specifications. The remaining discussions within this and the following chapters will assume the 15 kHz subcarrier spacing unless explicitly stated otherwise.

In the time domain, LTE transmissions are organized into *frames* of length 10 ms, each of which is divided into ten equally sized *subframes* of length 1 ms, as illustrated in Figure 9.1. Each subframe consists of two equally sized *slots* of length $T_{\text{slot}} = 0.5$ ms, with each slot consisting of a number of OFDM symbols including cyclic prefix.[1] To provide consistent and exact timing definitions, different time intervals within the LTE specifications are defined as multiples of a basic time unit $T_s = 1/(15000 \cdot 2048)$. The basic time unit T_s can thus be seen as

[1]This is valid for the "normal" downlink subframes. As described in Section 9.4, the *special subframe* present in case of TDD operation is not divided into two slots but rather into three fields.

Copyright Elsevier 2024

the sampling time of an FFT-based transmitter/receiver implementation with an FFT size equal to 2048. The time intervals outlined in Figure 9.1 can also be expressed as $T_{\text{frame}} = 307200 \cdot T_s$, $T_{\text{subframe}} = 30720 \cdot T_s$, and $T_{\text{slot}} = 15360 \cdot T_s$ for the frame, subframe, and slot durations respectively.

On a higher level, each frame is identified by a *System Frame Number* (SFN). The SFN is used to define different transmission cycles that have a period longer than one frame, such as paging sleep-mode cycles and periods for channel-state reporting. The SFN period equals 1024, thus the SFN repeats itself after 1024 frames or roughly 10 seconds.

The 15 kHz LTE subcarrier spacing corresponds to a useful symbol time $T_u = 2048 \cdot T_s$ or approximately 66.7 μs. The overall OFDM symbol time is then the sum of the useful symbol time and the cyclic-prefix length T_{CP}. As illustrated in Figure 9.1, LTE defines two cyclic-prefix lengths, the *normal* cyclic prefix and an *extended* cyclic prefix, corresponding to seven and six OFDM symbols per slot respectively. The exact cyclic-prefix lengths, expressed in the basic time unit T_s, are given in Figure 9.1. It can be noted that, in the case of the normal cyclic prefix, the cyclic-prefix length for the first OFDM symbol of a slot is somewhat larger compared to the remaining OFDM symbols. The reason for this is simply to fill the entire 0.5 ms slot, as the number of basic time units T_s per slot (15360) is not divisible by seven.

The reasons for defining two cyclic-prefix lengths for LTE are twofold:

- A longer cyclic prefix, although less efficient from a cyclic-prefix-overhead point of view, may be beneficial in specific environments with extensive delay spread, for example in very large cells. It is important to have in mind, though, that a longer cyclic prefix is not necessarily beneficial in the case of large cells, even if the delay spread is very extensive in such cases. If, in large cells, link performance is limited by noise rather than by signal corruption due to residual time dispersion not covered by the cyclic prefix, the additional robustness to radio-channel time dispersion, due to the use of a longer cyclic prefix, may not justify the corresponding additional energy overhead of a longer cyclic prefix.
- As already discussed in Chapter 3, in the case of MBSFN-based multicast/broadcast transmission, the cyclic prefix should not only cover the main part of the actual channel time dispersion, but also the timing difference between the transmissions received from the cells involved in the MBSFN transmission. In the case of MBSFN operation, the extended cyclic prefix is therefore used.

It should be noted that different cyclic-prefix lengths may be used for different subframes within a frame. As an example, to be discussed further in Chapter 17, MBSFN-based multicast/broadcast transmission is always confined to a limited set of subframes, in which case the use of the extended cyclic prefix, with its associated additional cyclic-prefix overhead, may only be applied to these subframes.[2]

[2]The extended cyclic prefix is then actually applied only to the so-called MBSFN part of the MBSFN subframes (see Section 9.2).

Copyright Elsevier 2024

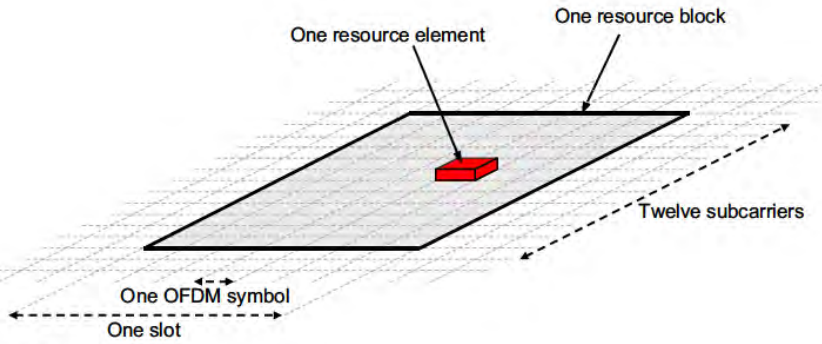


FIGURE 9.2

The LTE physical time-frequency resource

A *resource element*, consisting of one subcarrier during one OFDM symbol, is the smallest physical resource in LTE. Furthermore, as illustrated in Figure 9.2, resource elements are grouped into *resource blocks*, where each resource block consists of 12 consecutive subcarriers in the frequency domain and one 0.5 ms slot in the time domain. Each resource block thus consists of $7 \cdot 12 = 84$ resource elements in the case of a normal cyclic prefix and $6 \cdot 12 = 72$ resource elements in the case of an extended cyclic prefix.

Although resource blocks are defined over one slot, the basic time-domain unit for dynamic scheduling in LTE is one subframe, consisting of two consecutive slots. The reason for defining the resource blocks over one slot is that *distributed downlink transmission* (described in Chapter 10) and *uplink frequency hopping* (described in Chapter 11) are defined on a slot basis. The minimum scheduling unit, consisting of two time-consecutive resource blocks within one subframe (one resource block per slot), is referred to as a *resource-block pair.*

The LTE physical-layer specifications allow for a carrier to consist of any number of resource blocks in the frequency domain, ranging from a minimum of six resource blocks up to a maximum of 110 resource blocks. This corresponds to an overall transmission bandwidth ranging from roughly 1 MHz up to in the order of 20 MHz with very fine granularity and thus allows for a very high degree of LTE bandwidth flexibility, at least from a physical-layer-specification point of view. However, as mentioned in Chapter 7, LTE radio-frequency requirements are, at least initially, only specified for a limited set of transmission bandwidths, corresponding to a limited set of possible values for the number of resource blocks within a carrier. Also note that, in LTE release 10 and later, the total bandwidth of the transmitted signal can be as large as 100 MHz by aggregating multiple carriers, as mentioned in Chapter 7 and further described in Section 9.5.

Copyright Elsevier 2024

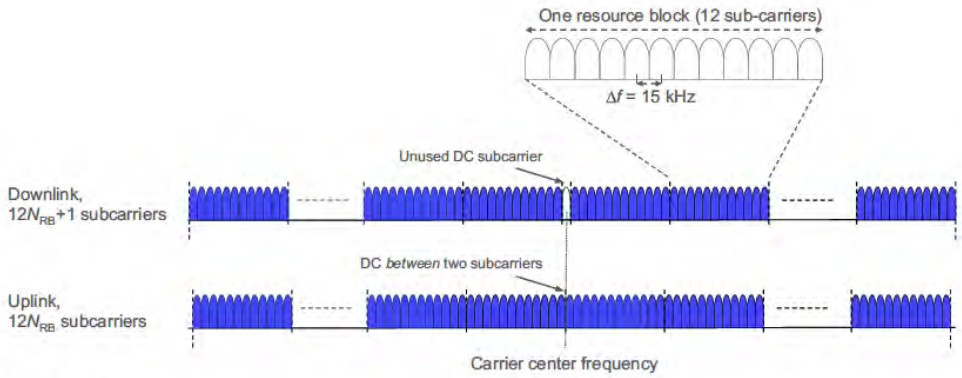


FIGURE 9.3

Frequency-domain structure for LTE

The resource-block definition above applies to both the downlink and uplink transmission directions. However, there is a minor difference between the downlink and uplink in terms of where the carrier center frequency is located in relation to the subcarriers.

In the downlink (upper part of Figure 9.3), there is an unused *DC-subcarrier* that coincides with the carrier center frequency. The reason why the DC-subcarrier is not used for downlink transmission is that it may be subject to disproportionately high interference due to, for example, local-oscillator leakage.

On the other hand, in the uplink (lower part of Figure 9.3), no unused DC-subcarrier is defined and the center frequency of an uplink carrier is located *between* two uplink subcarriers. The presence of an unused DC-carrier in the center of the spectrum would have prevented the assignment of the entire cell bandwidth to a single terminal and while still retaining the assumption of mapping to consecutive inputs of the OFDM modulator, something that is needed to retain the low-cubic-metric property of the DFTS-OFDM modulation used for uplink data transmission.

9.2 Normal subframes and MBSFN subframes

In LTE, each downlink subframe (and the DwPTS in the case of TDD; see Section 9.4.2 for a discussion of the TDD frame structure) is normally divided into a *control region*, consisting of the first few OFDM symbols, and a *data region*, consisting of the remaining part of the subframe. The usage of the resource elements in the two regions will be discussed in detail in Chapter 10; at this stage all we need to know is that the control region carries L1/L2 signaling necessary to control uplink and downlink data transmissions.

Copyright Elsevier 2024

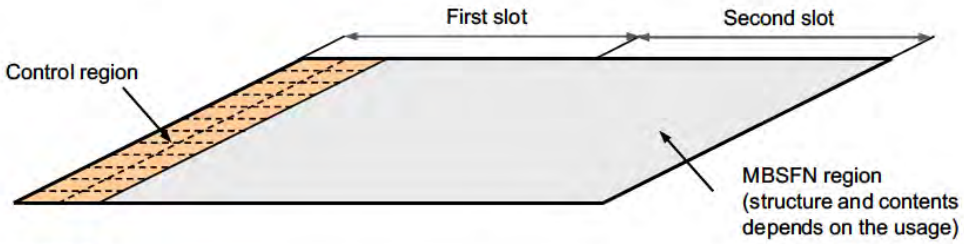


FIGURE 9.4

Resource-block structure for MBSFN subframes, assuming normal cyclic prefix for the control region and extended cyclic prefix for the MBSFN region

Additionally, from the first release of LTE so-called *MBSFN subframes* have been defined. The original intention with MBSFN subframes was, as indicated by the name, to support MBSFN transmission, as described in Chapter 17. However, MBSFN subframes have also been found to be useful in other contexts, for example as part of relaying functionality as discussed in Chapter 18. Hence, MBSFN subframes are therefore better seen as a generic tool and not related to MBSFN transmission only.

An MBSFN subframe, illustrated in Figure 9.4, consists of a control region of length one or two OFDM symbols, which is in essence identical to its counterpart in a normal subframe, followed by an *MBSFN region* whose contents depend on the usage of the MBSFN subframe. The reason for keeping the control region also in MBSFN subframes is, for example, to be able to transmit control signaling necessary for uplink transmissions. All terminals, from LTE release 8 and onwards, are capable of receiving the control region of an MBSFN subframe. This is the reason why MBSFN subframes have been found useful as a generic tool to introduce, in a backwards-compatible way, new types of signaling and transmission not part of an earlier release of the LTE radio-access specification. Such transmissions can be carried within the MBSFN region of the subframe and earlier-release terminals, not recognizing these transmissions, will simply ignore them.

Information about the set of subframes that are configured as MBSFN subframes in a cell is provided as part of the system information. In principle, an arbitrary pattern of MBSFN subframes can be configured with the pattern repeating after 40 ms.[3] However, as information necessary to operate the system (to be more specific, synchronization signals, system information, and paging, all of which will be discussed in detail in later chapters) needs to be transmitted in order for terminals to find and connect to a cell, subframes where such information is provided cannot be configured as MBSFN subframes. Therefore, subframes 0, 4, 5, and 9 for FDD and subframes 0, 1, 5, and 6 for TDD cannot be configured as

[3]The reason for 40 ms is that both the 10 ms frame length and the 8 ms hybrid-ARQ roundtrip-time are factors in 40 ms, which is important for many use cases, for example, relaying (see further Chapter 18).

Copyright Elsevier 2024

MBSFN subframes, leaving the remaining six subframes of a frame as candidates for MBSFN subframes.

9.3 Antenna ports

Downlink multi-antenna transmission is a key technology of LTE. Signals transmitted from different antennas or signals subject to different and, for the receiver, unknown *multi-antenna precoders* (see Section 10.3) will experience different "radio channels" even if the set of antennas are located at the same site.[4]

In general it is important for a terminal to understand what it can assume in terms of the relation between the radio channels experienced by different downlink transmissions. This is, for example, important in order for the terminal to be able to understand what reference signal(s) should be used for channel estimation for a certain downlink transmission. It is also important in order for the terminal to be able to determine relevant channel-state information, for example for scheduling and link-adaptation purposes.

For this reason, the concept of *antenna port* has been introduced in the LTE specifications. Each individual downlink transmission is carried out from a specific antenna port, the identity of which is known to the terminal. Furthermore, the terminal can assume that two transmitted signals have experienced the same radio channel *if and only if* they are transmitted from the same antenna port.[5]

In practice, each antenna port can, at least for the downlink, be seen as corresponding to a specific reference signal. A terminal receiver can then assume that this reference signal can be used to estimate the channel corresponding to the specific antenna port. The reference signals can also be used by the terminal to derive detailed channel-state information related to the antenna port.

It should be understood that an antenna port is an abstract concept that does not necessarily correspond to a specific physical antenna:

- Two different signals may be transmitted in the same way from multiple physical antennas. A terminal receiver will then see the two signals as propagating over a single channel corresponding to the "sum" of the channels of the different antennas, and the overall transmission could be seen as a transmission from a single antenna port being the same for the two signals.
- Two signals may be transmitted from the same set of antennas, but with different, for the receiver unknown, antenna transmitter-side precoders. A receiver will have to see the unknown antenna precoders as part of the overall channel, implying that the two signals will appear as having been transmitted from two different antenna ports. It should be

[4] An unknown transmitter-side precoder needs to be seen as part of the overall radio channel.
[5] For certain antenna ports, more specifically those that correspond to so-called demodulation reference signals, the assumption of the same radio channel is only valid within a given subframe.

Copyright Elsevier 2024

noted that if the antenna precoders of the two transmissions would have been known to be the same, the transmissions could have been seen as originating from the same antenna port. The same would have been true if the precoders would have been known to the receiver as, in that case, the precoders would not need to be seen as part of the radio channel.

9.3.1 Quasi-co-located antenna ports

Even if two signals have been transmitted from two different antennas, the channels experienced by the two signals may still have many *large-scale* properties in common. As an example, the channels experienced by two signals transmitted from two different antenna ports corresponding to different physical antennas at the same site will, even if being different in the details, typically have the same or at least similar large-scale properties, for example, in terms of Doppler spread/shift, average delay spread, and average gain. It can also be expected that the channels will introduce similar average delay. Knowing that the radio channels corresponding to two different antenna ports have similar large-scale properties can be used by the terminal receiver, for example, in the setting of parameters for channel estimation.

However, with the introduction of different types of multi-point transmission in LTE release 11, different downlink transmit antennas serving the same terminal may be much more geographically separated. In that case, the channels of different antenna ports relevant for a terminal may differ even in terms of large-scale properties.

For this reason the concept of *quasi-co-location* with respect to antenna ports was introduced as part of LTE release 11. A terminal receiver can assume that the radio channels corresponding to two different antenna ports have the same large-scale properties in terms of specific parameters such as average delay spread, Doppler spread/shift, average delay and average gain *if and only if* the antenna ports are specified as being quasi-co-located. Whether or not two specific antenna ports can be assumed to be quasi-co-located with respect to a certain channel property is in some cases given by the LTE specification. In other cases, the terminal may be explicitly informed by the network by means of signaling if two specific antenna ports can be assumed to be quasi-co-located or not.

As the name suggests, in practice, two antenna ports will typically be "quasi-co-located" if they correspond to physical antennas at the same location, while two antenna ports corresponding to antennas at different locations would typically not be "quasi-co-located." However, there is nothing explicitly stated in the specification about this, and quasi-co-location is simply defined with regards to what can be assumed regarding the relation between the long-term channel properties of different antenna ports.

9.4 Duplex schemes

Spectrum flexibility is one of the key features of LTE. In addition to the flexibility in transmission bandwidth, LTE also supports operation in both paired and unpaired spectrum

Copyright Elsevier 2024

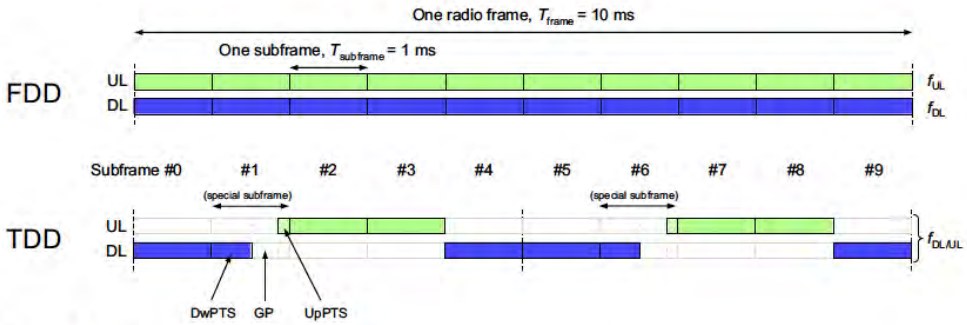


FIGURE 9.5

Uplink/downlink time-frequency structure in case of FDD and TDD

by supporting both FDD- and TDD-based duplex operation with the time–frequency structures illustrated in Figure 9.5. Although the time-domain structure is, in most respects, the same for FDD and TDD, there are some differences, most notably the presence of a *special subframe* in the case of TDD. The special subframe is used to provide the necessary guard time for downlink–uplink switching, as discussed below.

9.4.1 Frequency-division duplex (FDD)

In the case of FDD operation, uplink and downlink are carried on different carrier frequencies, denoted f_{UL} and f_{DL} in the upper part of Figure 9.5. During each frame, there are thus ten uplink subframes and ten downlink subframes, and uplink and downlink transmission can occur simultaneously within a cell. Isolation between downlink and uplink transmissions is achieved by transmission/reception filters, known as duplex filters, and a sufficiently large *duplex separation* in the frequency domain.

Even if uplink and downlink transmission can occur simultaneously within a cell in the case of FDD operation, a terminal may be capable of *full-duplex* operation or only *half-duplex* operation for a certain frequency band, depending on whether or not it is capable of simultaneous transmission/reception. In the case of full-duplex capability, transmission and reception may also occur simultaneously at a terminal, whereas a terminal capable of only half-duplex operation cannot transmit and receive simultaneously. As mentioned in Chapter 7, supporting only half-duplex operation allows for simplified terminal implementation due to relaxed duplex-filter requirements. This applies especially for certain frequency bands with a narrow duplex gap. Hence, full duplex support is *frequency-band dependent* such that a terminal may support only half-duplex operation in certain frequency bands while being capable of full-duplex operation in the remaining supported bands. It should be noted that full/half-duplex capability is a property of the *terminal*; the base station is operating in full duplex

Copyright Elsevier 2024

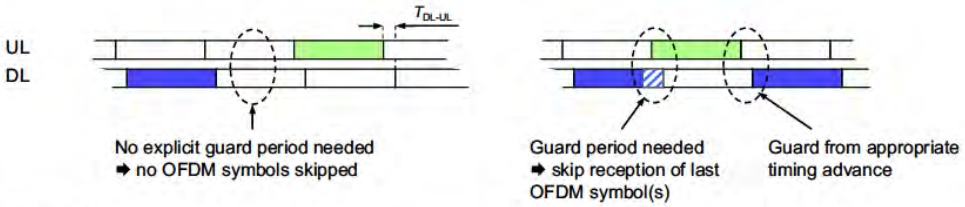


FIGURE 9.6

Guard time at the terminal for half-duplex FDD

irrespective of the terminal capabilities. As the relevant transmission structures and timing relations are identical between full-duplex and half-duplex FDD, a single cell may simultaneously support a mixture of full-duplex and half-duplex FDD terminals. Half-duplex operation has an impact on the sustained data rates that can be provided to/from a single mobile terminal as it cannot transmit in all uplink subframes, but the cell capacity is hardly affected as typically it is possible to schedule different terminals in uplink and downlink in a given subframe.

Since a half-duplex terminal is not capable of simultaneous transmission and reception, the scheduling decisions must take this into account and half-duplex operation can be seen as a scheduling restriction, as will be discussed in more detail in Chapter 13. If a terminal is scheduled such that downlink reception in one subframe immediately precedes a subframe of uplink transmission, a guard time is necessary for the terminal to switch from reception to transmission. This is created in such cases by allowing the terminal to skip receiving[6] the last OFDM symbol(s) in the downlink subframe, as illustrated in Figure 9.6.

Finally, it is worth noting that all FDD frequency bands specified by RAN4 so far (2013) are full-duplex-only bands—that is, it is assumed that, for these bands, all terminals support full-duplex operation.

9.4.2 Time-division duplex (TDD)

In the case of TDD operation (lower part of Figure 9.5), there is a single carrier frequency and uplink and downlink transmissions are separated in the time domain on a cell basis. As seen in the figure, some subframes are allocated for uplink transmissions, and some subframes are allocated for downlink transmission, with the switch between downlink and uplink occurring in a *special subframe* (subframe 1 and, for some downlink/uplink configurations, also subframe 6). Different asymmetries in terms of the amount of resources—that is, subframes—allocated for uplink and downlink transmission respectively are provided through the seven

[6]The impact on the decoding performance can be accounted for by the rate adaptation mechanism. For very high data rates, the performance impact may be somewhat larger due to the downlink interleaving structure, in which case it may be preferable not to schedule an uplink transmission immediately after a downlink reception.

Copyright Elsevier 2024

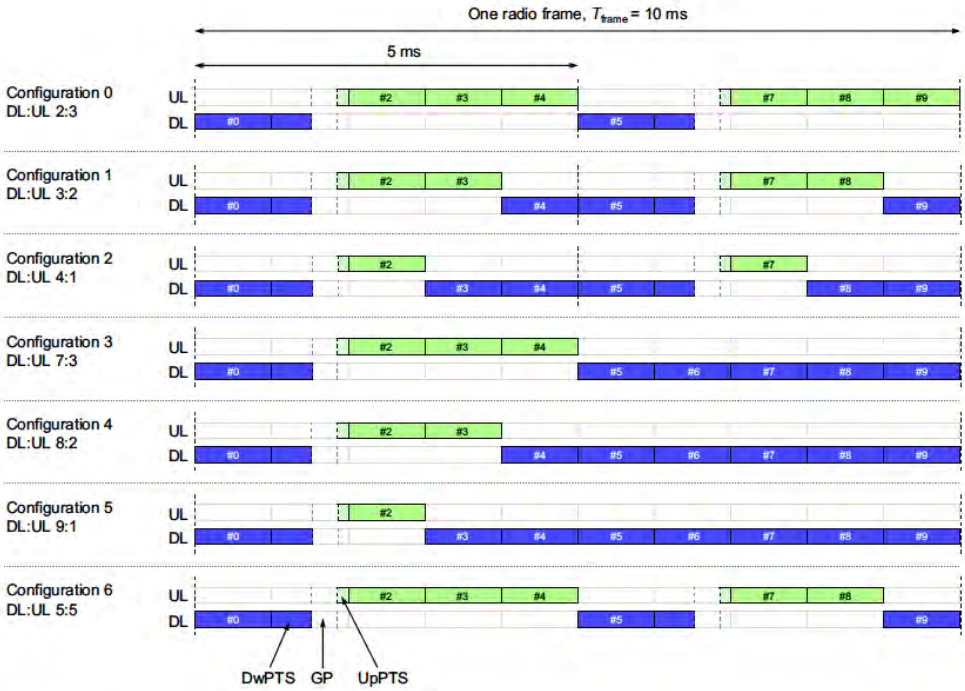


FIGURE 9.7

Different downlink/uplink configurations in case of TDD

different downlink/uplink configurations illustrated in Figure 9.7. As seen in the figure, subframes 0 and 5 are always allocated for downlink transmission while subframe 2 is always allocated for uplink transmissions. The remaining subframes (except the special subframe; see below) can then be flexibly allocated for downlink or uplink transmission depending on the downlink/uplink configuration. To avoid severe interference between downlink and uplink transmissions in different cells, neighboring cells typically have the same downlink/uplink configuration. This makes it difficult to change the downlink/uplink configuration dynamically, for example on a frame-by-frame basis. Therefore, the current LTE specification assumes that the downlink/uplink configuration is relatively static. It can be changed on a very slow basis though, in order to adapt to changing traffic patterns. It could also, in principle, be different in different areas, for example to match different traffic patterns, although inter-cell interference needs to be carefully addressed in this case.

As the same carrier frequency is used for uplink and downlink transmission, both the base station and the terminal need to switch from transmission to reception and vice versa. The

Copyright Elsevier 2024

switch between downlink and uplink occurs in the special subframe, which is split into three parts: a downlink part (DwPTS), a guard period (GP), and an uplink part (UpPTS).

The DwPTS is in essence treated as a normal downlink subframe,[7] although the amount of data that can be transmitted is smaller due to the reduced length of the DwPTS compared to a normal subframe. The UpPTS, however, is not used for data transmission due to the very short duration. Instead, it can be used for channel sounding or random access. It can also be left empty, in which case it serves as an extra guard period.

An essential aspect of any TDD system is the possibility to provide a sufficiently large *guard period* (or guard time), where neither downlink nor uplink transmissions occur. This guard period is necessary for switching from downlink to uplink transmission and vice versa and, as already mentioned, is obtained from the special subframe. The required length of the guard period depends on several factors. First, it should be sufficiently large to provide the necessary time for the circuitry in base stations and the terminals to switch from downlink to uplink. Switching is typically relatively fast, of the order of 20 microseconds, and in most deployments does not significantly contribute to the required guard time.

Secondly, the guard time should also ensure that uplink and downlink transmissions do not interfere at the base station. This is handled by advancing the uplink timing at the terminals such that, at the base station, the last uplink subframe before the uplink-to-downlink switch ends before the start of the first downlink subframe. The uplink timing of each terminal can be controlled by the base station by using the timing advance mechanism, as will be elaborated upon in Chapter 11. The guard period must be large enough to allow the terminal to receive the downlink transmission and switch from reception to transmission before it starts the (timing-advanced) uplink transmission. In essence, some of the guard period of the special subframe is "moved" from the downlink-to-uplink switch to the uplink-to-downlink switch by the timing-advance mechanism. This is illustrated in Figure 9.8. As the timing advance is proportional to the distance to the base station, a larger guard period is required when operating in large cells compared to small cells.

Finally, the selection of the guard period also needs to take interference between base stations into account. In a multi-cell network, inter-cell interference from downlink transmissions in neighboring cells must decay to a sufficiently low level before the base station can start to receive uplink transmissions. Hence, a larger guard period than is motivated by the cell size itself may be required as the last part of the downlink transmissions from distant base stations otherwise may interfere with uplink reception. The amount of guard period depends on the propagation environments, but in some cases the inter-base-station interference is a non-negligible factor when determining the guard period.

From the above discussion, it is clear that a sufficient amount of configurability of the guard period is needed to meet different deployment scenarios. Therefore, a set of DwPTS/GP/UpPTS configurations is supported as shown in Table 9.1, where each configuration

[7]For the shortest DwPTS duration of 3 OFDM symbols, DwPTS cannot be used for PDSCH transmission.

Copyright Elsevier 2024

Samsung Exhibit 1017, Page 178 of 538
Samsung Electronics Co., Ltd. v. Asus Technology Licensing Inc.
IPR2024-01217

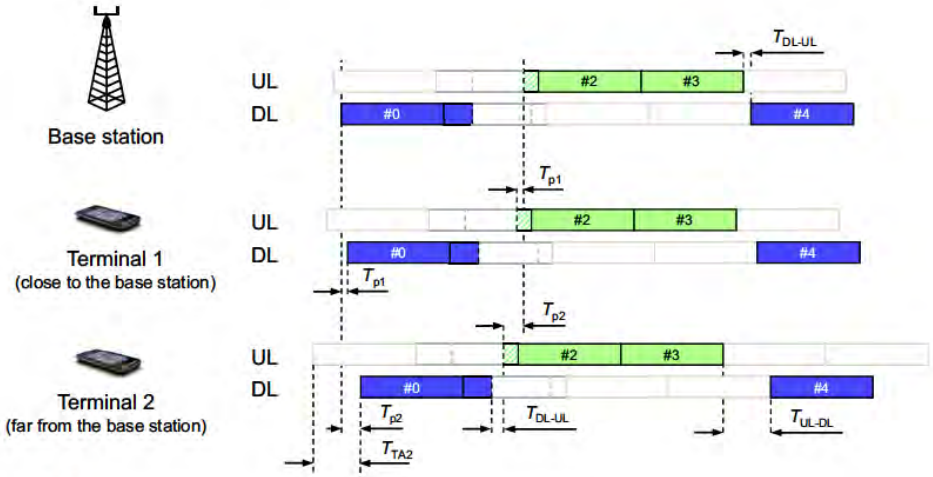


FIGURE 9.8

Timing relation for TDD operation

Table 9.1 Resulting Guard Period in OFDM Symbols for Different DwPTS and UpPTS Lengths (Normal Cyclic Prefix)

	DwPTS									
	12	11		10		9		6[a]	3	
GP	1	1	2	2	3	3	4	6	9	10
UpPTS	1	2	1	2	1	2	1	2	2	1

[a]*The 6:6:2 configuration was added in release 11 to improve efficiency when coexisting with some common TD-SCDMA configurations (terminals prior to release 11 use 3:9:2).*

corresponds to a given length of the three fields in the special subframes. The DwPTS/GP/UpPTS configuration used in the cell is signaled as part of the system information.

9.4.3 LTE and TD-SCDMA coexistence

In addition to supporting a wide range of different guard periods, an important aspect in the design of TDD in LTE was to simplify coexistence with and migration from systems based on the 3GPP TD-SCDMA standard.[8] Basically, to handle inter-system interference from two different but co-sited TDD systems operating close in frequency, it is necessary to align

[8]TD-SCDMA is one of three TDD modes defined by 3GPP for UTRA TDD and the only one deployed on a larger scale.

Copyright Elsevier 2024

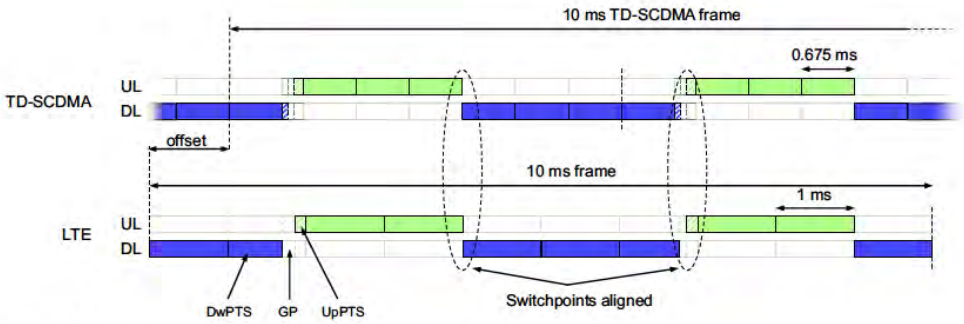


FIGURE 9.9

Coexistence between TD-SCDMA and LTE

the switch-points between the two systems. Since LTE supports configurable lengths of the DwPTS field, the switch-points of LTE and TD-SCDMA (or any other TDD system) can be aligned, despite the different subframe lengths used in the two systems. Aligning the switch-points between TD-SCDMA and LTE is the technical reason for splitting the special subframe into the three fields DwPTS/GP/UpPTS instead of locating the switch-point at the subframe boundary. An example of LTE/TD-SCDMA coexistence is given in Figure 9.9.

The set of possible lengths of DwPTS/GP/UpPTS is selected to support common coexistence scenarios, as well as to provide a high degree of guard-period flexibility for the reasons discussed earlier. The UpPTS length is one or two OFDM symbols and the DwPTS length can vary from three[9] to twelve OFDM symbols, resulting in guard periods ranging from one to ten OFDM symbols. The resulting guard period for the different DwPTS and UpPTS configurations supported is summarized in Table 9.1 for the case of normal cyclic prefix. As discussed earlier, the DwPTS can be used for downlink data transmission, while the UpPTS can be used for sounding or random access only, due to its short duration.

9.5 Carrier aggregation

The possibility for *carrier aggregation* was introduced in LTE release 10. In the case of carrier aggregation, multiple LTE carriers, each with a bandwidth up to 20 MHz, can be transmitted in parallel to/from the same terminal, thereby allowing for an overall wider bandwidth and correspondingly higher per-link data rates. In the context of carrier

[9]The smallest DwPTS length is motivated by the location of the primary synchronization signal in the DwPTS (see Chapter 14).

Copyright Elsevier 2024

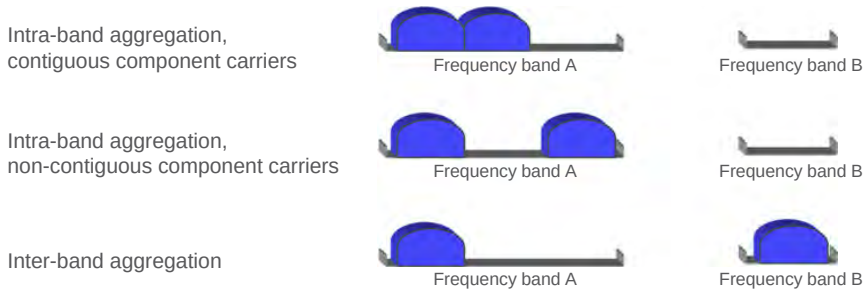


FIGURE 9.10

Different types of carrier aggregation

aggregation, each carrier is referred to as a *component carrier*[10] as, from an RF point-of-view, the entire set of aggregated carriers can be seen as a single (RF) carrier.

Up to five component carriers, possibly of different bandwidths up to 20 MHz, can be aggregated allowing for overall transmission bandwidths up to 100 MHz. A terminal capable of carrier aggregation may receive or transmit simultaneously on multiple component carriers. Each component carrier can also be accessed by an LTE terminal from earlier releases—that is, component carriers are *backwards compatible*. Thus, in most respects and unless otherwise mentioned, the physical-layer description in the following chapters applies to each component carrier separately in the case of carrier aggregation.

It should be noted that aggregated component carriers do not need to be contiguous in the frequency domain. Rather, with respect to the frequency location of the different component carriers, three different cases can be identified (see also Figure 9.10):

- Intra-band aggregation with frequency-contiguous component carriers.
- Intra-band aggregation with non-contiguous component carriers.
- Inter-band aggregation with non-contiguous component carriers.

The possibility to aggregate non-adjacent component carriers allows for exploitation of a fragmented spectrum; operators with a fragmented spectrum can provide high-data-rate services based on the availability of a wide overall bandwidth, even though they do not possess a single wideband spectrum allocation. Except from an RF point of view there is no difference between the three different cases outlined in Figure 9.10 and they are all supported by the basic LTE release-10 specification. However, the complexity of RF implementation is vastly different (as discussed in Chapter 19), with the first case being the least complex. Thus, although spectrum aggregation is supported by the physical-layer and protocol specifications, the actual implementation will be strongly constrained, including specification of only a

[10]In the specifications, the term cell is used instead of component carrier, but as the term cell is something of a misnomer in the uplink case, the term component carrier is used herein.

Copyright Elsevier 2024

Samsung Exhibit 1017, Page 181 of 538
Samsung Electronics Co., Ltd. v. Asus Technology Licensing Inc.
IPR2024-01217

limited number of aggregation scenarios and aggregation over a dispersed spectrum only being supported by the most advanced terminals.

Although carrier aggregation can be utilized to achieve very high overall bandwidths, up to 100 MHz, very few operators, if any, have such large spectrum allocations. Rather, the main usage of carrier aggregation, at least in the short-term perspective, is to handle fragmented spectrum allocations in which an operator may have 5–10 MHz of spectrum allocation in several bands and, despite this, would like to provide end-user performance on par with an operator having a larger amount of contiguous spectrum.

A terminal capable of carrier aggregation has one downlink *primary* component carrier and an associated uplink primary component carrier. In addition, it may have one or several *secondary* component carriers in each direction. Different terminals may have different carriers as their primary component carrier—that is, the configuration of the primary component carrier is terminal specific. The association between the downlink primary carrier and the corresponding uplink primary carrier is cell specific and signaled as part of the system information. This is similar to the case without carrier aggregation, although in the latter case the association is trivial. The reason for such an association is, for example, to determine to which uplink component carrier a certain scheduling grant transmitted on the downlink relates, without having to explicitly signal the component-carrier number.

All idle mode procedures apply to the primary component carrier only, or, expressed differently, carrier aggregation with additional secondary carriers configured only applies to terminals in the RRC_CONNECTED state. Upon connection to the network, the terminal performs the related procedures such as cell search and random access (see Chapter 14 for a detailed description of these procedures) following the same steps as in the absence of carrier aggregation. Once the communication between the network and the terminal is established, additional secondary component carriers can be configured.

The fact that carrier aggregation is terminal specific—that is, different terminals may be configured to use different sets of component carriers—is useful not only from a network perspective to balance the load across component carriers, but also to handle different capabilities between terminals. Some terminals may be able to transmit/receive on multiple component carriers, while other terminals may do so on only a single carrier. This is a consequence of being able to serve terminals from earlier releases at the same time as a carrier-aggregation-capable release-10 terminal, but it also allows for different capabilities in terms of carrier aggregation for different terminals as well as a differentiation between downlink and uplink carrier-aggregation capability. For example, a terminal may be capable of two component carriers in the downlink but of only a single component carrier—that is, no carrier aggregation—in the uplink, as is the case for terminal C in Figure 9.11. Note also that the primary component-carrier configuration can differ between terminals. Asymmetric carrier aggregation can also be useful to handle different spectrum allocations, for example if an operator has more spectrum available for downlink transmissions than uplink transmissions.

Copyright Elsevier 2024

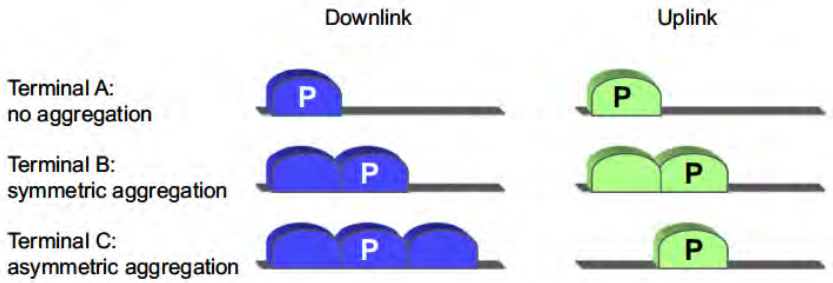


FIGURE 9.11

Examples of carrier aggregation ("P" denotes the primary component carrier)

In releases 10 and 11, only downlink-heavy asymmetries are supported—that is, the number of uplink component carriers configured for a terminal is always equal to or smaller than the number of configured downlink component carriers. Uplink-heavy asymmetries are less likely to be of practical interest and would also complicate the overall control signaling structure, as in such a case multiple uplink component carriers would need to be associated with the same downlink component carrier.

Carrier aggregation is supported for both FDD and TDD, although all component carriers need to have the same duplex scheme. Furthermore, in the case of TDD, the uplink–downlink configuration should be the same across component carriers in release 10, a restriction that was (partially) lifted in release 11.

For carrier-aggregation capable TDD terminals in release 11, different uplink-downlink configurations can be used for component carriers in *different* frequency bands. The main motivation is improved support for coexistence with other systems. One example is two independent legacy systems operating in two different frequency bands. If LTE is to exploit spectrum in both of these bands through carrier aggregation, the uplink-downlink allocation of the component carriers in the respective band is basically determined by the legacy systems and may very well be different.

It is worth noting that inter-band carrier aggregation with *different* uplink-downlink configurations may imply *simultaneous* uplink transmission and downlink reception in the terminal. Terminals capable of simultaneous transmission and reception need, similarly to FDD terminals and unlike most TDD terminals, a duplex filter. Whether a TDD terminal is equipped with a duplex filter and capable of simultaneous transmission and reception is therefore a terminal capability. Terminals not having this capability follow the uplink-downlink configuration on the primary carrier. They do not transmit in the uplink on a secondary component carrier whenever there is a downlink subframe on the primary component carrier (and vice versa). In essence, this implies that certain subframes on the secondary component carriers are not usable by terminals not capable of simultaneous reception and transmission. This is illustrated in Figure 9.12 (the downlink subframe on the secondary

Copyright Elsevier 2024

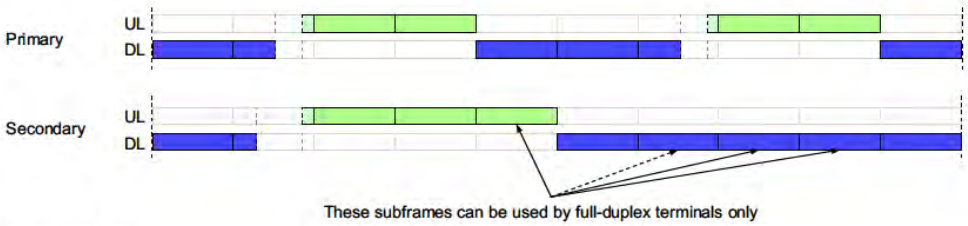


FIGURE 9.12

Examples of TDD inter-band carrier aggregation with different downlink-uplink configurations

component carrier overlapping with the special subframe on the primary component carrier can only be used in the part overlapping with the DwPTS).

The special subframe configuration can be different for the different component carriers although release 10 and release 11 terminals not capable of simultaneous transmission and reception require that the resulting downlink–uplink switch time is sufficiently large.

9.6 Frequency-domain location of LTE carriers

In principle, an LTE carrier could be positioned anywhere within the spectrum and, actually, the basic LTE physical-layer specification does not say anything about the exact frequency location of an LTE carrier, including the frequency band. However, in practice, there is a need for restrictions on where an LTE carrier can be positioned in the frequency domain.

- In the end, an LTE terminal must be implemented and RF-wise such a terminal can only support certain frequency bands. The frequency bands for which LTE is specified to operate are discussed in Chapter 19.
- After being activated, an LTE terminal has to search for a network-transmitted carrier within the frequency bands supported by the terminal. In order for that carrier search to not take an unreasonably long time, there is a need to limit the set of frequencies to be searched.

For this reason it is assumed that, within each supported frequency band, LTE carriers may exist on a 100 kHz carrier raster or carrier grid—that is, the carrier center frequency can be expressed as $m \cdot 100$ kHz, where m is an integer (see Figure 9.13).

In the case of carrier aggregation, multiple carriers can be transmitted to/from the same terminal. In order for the different component carriers to be accessible by earlier-release terminals, each component carrier should fall on the 100 kHz carrier grid. However, there is an additional constraint that the carrier spacing between adjacent component carriers should be a multiple of the 15 kHz subcarrier spacing to allow transmission/reception with a

Copyright Elsevier 2024

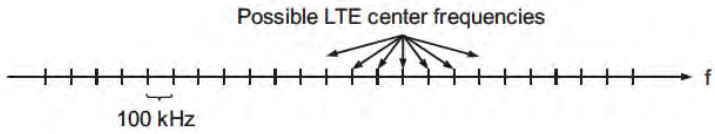


FIGURE 9.13

LTE carrier raster

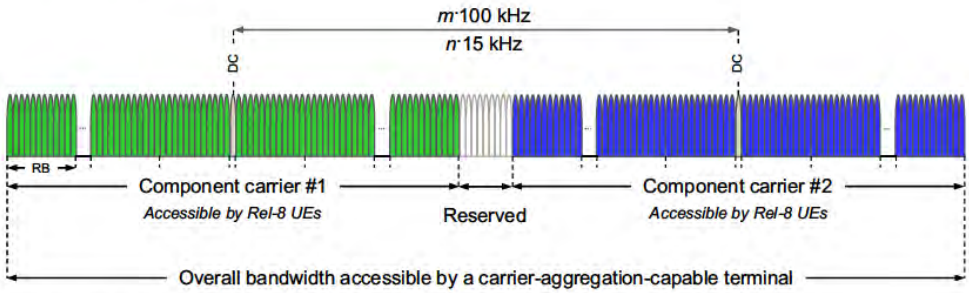


FIGURE 9.14

LTE carrier raster and carrier aggregation

single FFT.[11] Thus, in the case of carrier aggregation, the carrier spacing between the different component carriers should be a multiple of 300 kHz, the smallest carrier spacing being a multiple of both 100 kHz (the raster grid) and 15 kHz (the subcarrier spacing). A consequence of this is that there will always be a small gap between two component carriers, even when they are located as close as possible to each other, as illustrated in Figure 9.14.

[11]This is only relevant for component carriers that are contiguous in the frequency domain. Furthermore, in case of independent frequency errors between component carriers, separate FFTs may be needed at the receiver.

Copyright Elsevier 2024

This page intentionally left blank

Copyright Elsevier 202

CHAPTER

Downlink Physical-Layer Processing

10

In Chapter 8, the LTE radio-interface architecture was discussed with an overview of the functions and characteristics of the different protocol layers. Chapter 9 then gave an overview of the time-frequency structure of LTE transmissions including the structure of the basic OFDM time-frequency grid being the fundamental physical resource on both uplink and downlink. It also discussed the concept of antenna ports, especially relevant for the understanding of multi-antenna and multi-point transmissions within LTE.

This chapter will provide a more detailed description of the downlink physical-layer functionality including the transport-channel processing (Section 10.1), reference signals (Section 10.2), multi-antenna transmission (Section 10.3) and L1/L2 control signaling (Section 10.4). Chapter 11 will provide a corresponding description for the *uplink* transmission direction. The later chapters will then go further into the details of some specific uplink and downlink functions and procedures.

10.1 Transport-channel processing

As described in Chapter 8, the physical layer provides services to the MAC layer in the form of transport channels. As also described, for the LTE downlink there are four different types of transport channels defined: the Downlink Shared Channel (DL-SCH), the Multicast Channel (MCH), the Paging Channel (PCH), and the Broadcast Channel (BCH). This section provides a detailed description of the physical-layer processing applied to the DL-SCH, including the mapping to the physical resource—that is, to the resource elements of the OFDM time–frequency grid of the set of antenna ports to be used for the transmission. DL-SCH is the transport-channel type in LTE used for transmission of downlink user-specific higher-layer information, both user data and dedicated control information, as well as the main part of the downlink system information (see Chapter 14). The DL-SCH physical-layer processing is to a large extent applicable also to MCH and PCH transport channels, although with some additional constraints. On the other hand, as mentioned in Chapter 8, the physical-layer processing, and the structure in general, for BCH transmission

4G: LTE/LTE-Advanced for Mobile Broadband. http://dx.doi.org/10.1016/B978-0-12-419985-9.00010-6
Copyright © 2014 Erik Dahlman, Stefan Parkvall and Johan Sköld. Published by Elsevier Ltd. All rights reserved.

Copyright Elsevier 2024

Samsung Exhibit 1017, Page 187 of 538
Samsung Electronics Co., Ltd. v. Asus Technology Licensing Inc.
IPR2024-01217

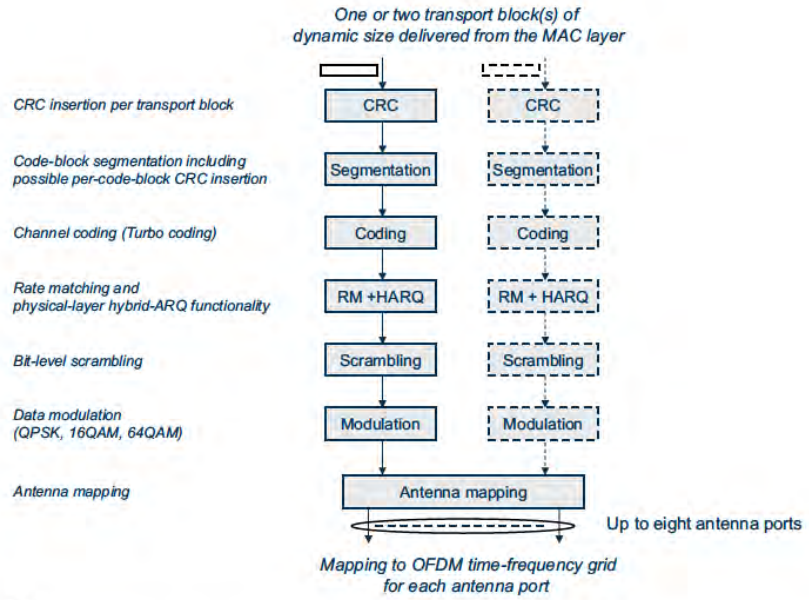


FIGURE 10.1

Physical-layer processing for the downlink shared channel (DL-SCH)

is quite different. BCH transmission is described in Chapter 14 as part of the discussion on LTE system information.

10.1.1 Processing steps

The different steps of the DL-SCH physical-layer processing are outlined in Figure 10.1. In the case of carrier aggregation—transmission on multiple component carriers in parallel to the same terminal—the transmissions on the different carriers correspond to separate transport channels with separate and essentially independent physical-layer processing. The transport-channel processing outlined in Figure 10.1 and the discussion below is thus valid also in the case of carrier aggregation.

Within each *Transmission Time Interval* (TTI), corresponding to one subframe of length one ms, up to two transport blocks of dynamic size are delivered to the physical layer and transmitted over the radio interface for each component carrier. The number of transport blocks transmitted within a TTI depends on the configuration of the multi-antenna transmission scheme (see Section 10.3).

Copyright Elsevier 2024

- In the case of no *spatial multiplexing* there is at most a single transport block in a TTI.
- In the case of spatial multiplexing, with transmission on multiple *layers* in parallel to the same terminal, there are two transport blocks within a TTI.[1]

10.1.1.1 CRC insertion per transport block

In the first step of the physical-layer processing, a 24-bit CRC is calculated for and appended to each transport block. The CRC allows for receiver-side detection of errors in the decoded transport block. The corresponding error indication can, for example, be used by the downlink hybrid-ARQ protocol as a trigger for requesting retransmissions.

10.1.1.2 Code-block segmentation and per-code-block CRC insertion

The LTE Turbo-coder internal interleaver is only defined for a limited number of code-block sizes, with a maximum block size of 6144 bits. If the transport block, including the transport-block CRC, exceeds this maximum code-block size, *code-block segmentation* is applied before the Turbo coding, as illustrated in Figure 10.2. Code-block segmentation implies that the transport block is segmented into smaller *code blocks*, the sizes of which should match the set of code-block sizes supported by the Turbo coder.

In order to ensure that a transport block of arbitrary size can be segmented into code blocks that match the set of available code-block sizes, the LTE specification includes the possibility to insert "dummy" *filler bits* at the head of the first code block. However, the set of transport-block sizes currently defined for LTE has been selected so that filler bits are not needed.

As can be seen in Figure 10.2, code-block segmentation also implies that an additional CRC (also of length 24 bits but different compared to the transport-block CRC described

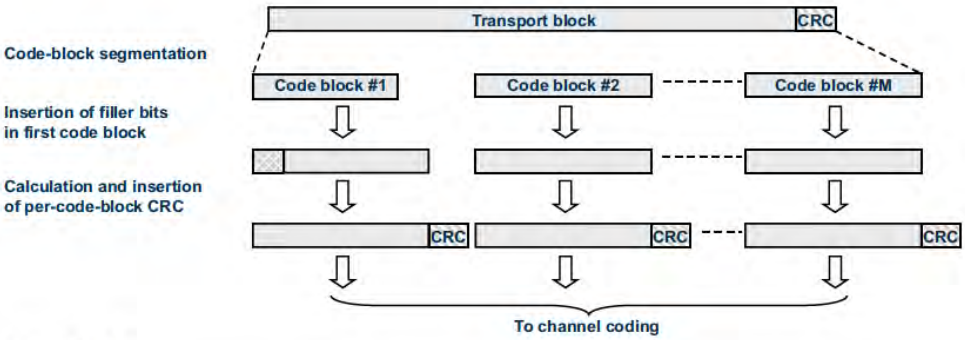


FIGURE 10.2

Code-block segmentation and per-code-block CRC insertion

[1]This is true for initial transmissions. In the case of hybrid-ARQ retransmissions there may also be cases when a single transport block is transmitted over multiple layers as discussed, for example, in Section 10.3.3.

Copyright Elsevier 2024

above) is calculated for and appended to each code block. Having a CRC per code block allows for early detection of correctly decoded code blocks and correspondingly early termination of the iterative decoding of that code block. This can be used to reduce the terminal processing effort and corresponding energy consumption. In the case of a single code block no additional code-block CRC is applied.

One could argue that, in case of code-block segmentation, the transport-block CRC is redundant and implies unnecessary overhead as the set of code-block CRCs should indirectly provide information about the correctness of the complete transport block. However, code-block segmentation is only applied to large transport blocks for which the relative extra overhead due to the additional transport-block CRC is small. The transport-block CRC also adds additional error-detection capabilities and thus further reduces the risk for undetected errors in the decoded transport block.

Information about the transport-block size is provided to the terminal as part of the scheduling assignment transmitted on the PDCCH/EPDCCH control channels, as described in Section 10.4. Based on this information, the terminal can determine the code-block size and number of code blocks. The terminal receiver can thus, based on the information provided in the scheduling assignment, straightforwardly undo the code-block segmentation and recover the decoded transport blocks.

10.1.1.3 Channel coding

Channel coding for DL-SCH (as well as for PCH and MCH) is based on Turbo coding [53], with encoding according to Figure 10.3. The encoding consists of two rate-1/2, eight-state constituent encoders, implying an overall code rate of 1/3, in combination with QPP-based[2]

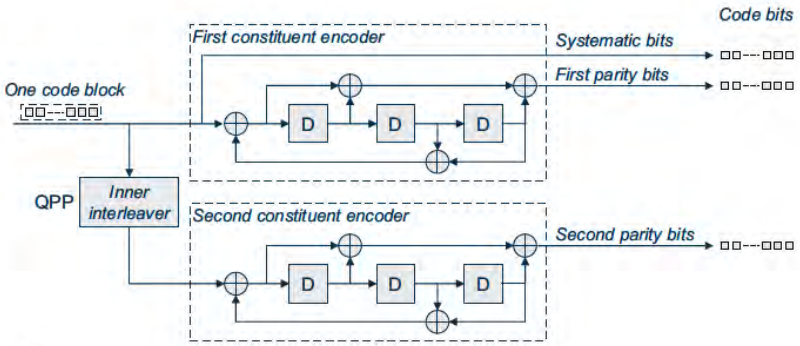


FIGURE 10.3

LTE Turbo encoder

[2]QPP = Quadrature Permutation Polynomial.

Copyright Elsevier 2024

Samsung Exhibit 1017, Page 190 of 538
Samsung Electronics Co., Ltd. v. Asus Technology Licensing Inc.
IPR2024-01217

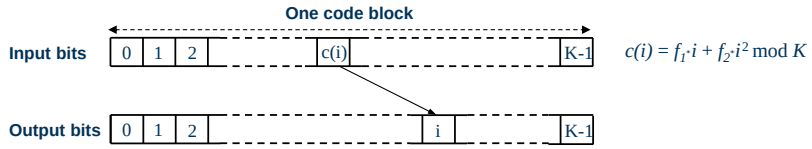


FIGURE 10.4

Principles of QPP-based interleaving

interleaving [69]. As illustrated in Figure 10.4, the QPP interleaver provides a mapping from the input (non-interleaved) bits to the output (interleaved) bits according to the function:

$$c(i) = f_1 \cdot i + f_2 \cdot i^2 \bmod K$$

where i is the index of the bit at the output of the interleaver, $c(i)$ is the index of the same bit at the input of the interleaver, and K is the code-block/interleaver size. The values of the parameters f_1 and f_2 depend on the code-block size K. The LTE specification lists all supported code-block sizes, ranging from a minimum of 40 bits to a maximum of 6144 bits, together with the associated values for the parameters f_1 and f_2. Thus, once the code-block size is known, the Turbo-coder inner interleaving, as well as the corresponding de-interleaving at the receiver side, can straightforwardly be carried out.

A QPP-based interleaver is *maximum contention free* [70], implying that the decoding can be parallelized without the risk for contention when the different parallel processes are accessing the interleaver memory. For the very high data rates supported by LTE, the improved possibilities for parallel processing offered by QPP-based interleaving can substantially simplify the Turbo-encoder/decoder implementation.

10.1.1.4 Rate matching and physical-layer hybrid-ARQ functionality

The task of the rate-matching and physical-layer hybrid-ARQ functionality is to extract, from the blocks of code bits delivered by the channel encoder, the exact set of code bits to be transmitted within a given TTI/subframe.

As illustrated in Figure 10.5, the outputs of the Turbo encoder (systematic bits, first parity bits, and second parity bits) are first separately interleaved. The interleaved bits are then inserted into what can be described as a circular buffer with the systematic bits inserted first, followed by alternating insertion of the first and second parity bits.

The bit selection then extracts consecutive bits from the circular buffer to an extent that matches the number of available resource elements in the resource blocks assigned for the transmission. The exact set of bits to extract depends on the *redundancy version* (RV) corresponding to different starting points for the extraction of coded bits from the circular buffer. As can be seen, there are four different alternatives for the redundancy version. The transmitter/scheduler selects the redundancy version and provides information about the selection as part of the scheduling assignment (see Section 10.4.6).

Copyright Elsevier 2024

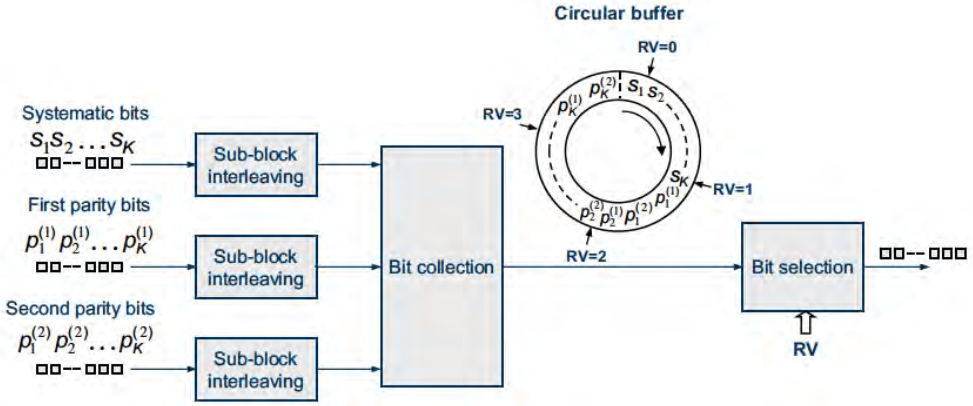


FIGURE 10.5

Rate-matching and hybrid-ARQ functionality

Note that the rate-matching and hybrid-ARQ functionality operates on the full set of code bits corresponding to one transport block and not separately on the code bits corresponding to a single code block.

10.1.1.5 Bit-level scrambling

LTE downlink scrambling implies that the block of code bits delivered by the hybrid-ARQ functionality is multiplied (*exclusive-or* operation) by a bit-level *scrambling sequence*. Without downlink scrambling, the channel decoder at the terminal could, at least in principle, be equally matched to an interfering signal as to the target signal, thus being unable to properly suppress the interference. By applying different scrambling sequences for neighboring cells, the interfering signal(s) after descrambling is (are) randomized, ensuring full utilization of the processing gain provided by the channel code. Thus, the bit scrambling essentially serves the same purpose as the scrambling applied at chip level after the direct-sequence spreading in DS-CDMA-based systems such as WCDMA/HSPA. Fundamentally, channel coding can be seen as "advanced" spreading providing processing gain similar to direct-sequence spreading but also additional coding gain.

In LTE, downlink scrambling is applied to all transport channels as well as to the downlink L1/L2 control signaling. For all downlink transport-channel types except MCH the scrambling sequences differ between neighboring cells (*cell-specific scrambling*). The scrambling also depends on the identity of the terminal to which the transmission is intended, assuming that the data is intended for a specific terminal. In contrast, in the case of MBSFN-based transmission using MCH, the same scrambling should be applied to all cells taking part in the MBSFN transmission—that is, all cells within the so-called *MBSFN area* (see Chapter 17). Thus, in the case of MCH transmission the scrambling depends on the MBSFN area identity.

Copyright Elsevier 2024

10.1.1.6 Data modulation

The downlink data modulation transforms the block of scrambled bits to a corresponding block of complex modulation symbols. The set of modulation schemes supported for the LTE downlink includes QPSK, 16QAM, and 64QAM, corresponding to two, four, and six bits per modulation symbol respectively.

10.1.1.7 Antenna mapping

The antenna mapping jointly processes the modulation symbols corresponding to the one or two transport blocks and maps the result to the set of antenna ports to be used for the transmission. The antenna mapping can be configured in different ways corresponding to different multi-antenna transmission schemes, including transmit diversity, beamforming, and spatial multiplexing. As indicated in Figure 10.1, LTE supports transmission using up to eight antenna ports depending on the exact multi-antenna transmission scheme. More details about LTE downlink multi-antenna transmission are provided in Section 10.3.

10.1.1.8 Resource-block mapping

The resource-block mapping takes the symbols to be transmitted on each antenna port and maps them to the set of available resource elements in the set of resource blocks assigned by the MAC scheduler for the transmission. As described in Chapter 9, each resource block consists of 84 resource elements (twelve subcarriers during seven OFDM symbols).[3] However, some of the resource elements within a resource block will not be available for the transport-channel transmission as they are occupied by:

- different types of downlink reference signals, as described in Section 10.2;
- downlink L1/L2 control signaling (one, two, or three OFDM symbols at the head of each subframe) as described in Section 10.4.[4]

Furthermore, as will be described in Chapter 14, within some resource blocks additional resource elements are reserved for the transmission of *synchronization signals* as well as for the PBCH physical channel which carries the BCH transport channel.

It should also now be pointed out that for the so-called Transmission Mode 10, the possibility for more dynamic control of the PDSCH mapping has been introduced to support multi-point transmission. This will be further discussed in Chapter 15 as part of the description of CoMP-related features introduced in LTE release 11.

10.1.2 Localized and distributed resource mapping

As discussed in Chapter 7, when deciding what set of resource blocks to use for transmission to a specific terminal, the network may take the downlink channel conditions in both the time

[3] 72 resource elements in case of extended cyclic prefix.
[4] In MBSFN subframes the control region is limited to a maximum of two OFDM symbols.

Copyright Elsevier 2024

and frequency domains into account. Such time/frequency-domain channel-dependent scheduling, taking channel variations—for example, due to frequency-selective fading—into account, may significantly improve system performance in terms of achievable data rates and overall cell throughput.

However, in some cases downlink channel-dependent scheduling is not suitable to use or is not practically possible:

- For low-rate services such as voice, the feedback signaling associated with channel-dependent scheduling may lead to extensive relative overhead.
- At high mobility (high terminal speed), it may be difficult or even practically impossible to track the instantaneous channel conditions to the accuracy required for channel-dependent scheduling to be efficient.

In such situations, an alternative means to handle radio-channel frequency selectivity is to achieve frequency diversity by distributing a downlink transmission in the frequency domain.

One way to distribute a downlink transmission in the frequency domain, and thereby achieve frequency diversity, is to assign multiple non-frequency-contiguous resource blocks for the transmission to a terminal. LTE allows for such *distributed resource-block allocation* by means of *resource allocation types 0 and 1* (see Section 10.4.6.1). However, although sufficient in many cases, distributed resource-block allocation by means of these resource-allocation types has certain drawbacks:

- For both types of resource allocations, the minimum size of the allocated resource can be as large as four resource-block pairs and may thus not be suitable when resource allocations of smaller sizes are needed.
- In general, both these resource-allocation methods are associated with a relatively large control-signaling overhead for the scheduling assignment (see Section 10.4.6).

In contrast, *resource-allocation type 2* (Section 10.4.6.1) always allows for the allocation of a single resource-block pair and is also associated with a relatively small control-signaling overhead. However, resource allocation type 2 only allows for the allocation of resource blocks that are contiguous in the frequency domain. In addition, regardless of the type of resource allocation, frequency diversity by means of distributed resource-block allocation will only be achieved in the case of resource allocations larger than one resource-block pair.

In order to provide the possibility for distributed resource-block allocation in the case of resource-allocation type 2, as well as to allow for distributing the transmission of a single resource-block pair in the frequency domain, the notion of a *Virtual Resource Block* (VRB) has been introduced for LTE.

What is being provided in the resource allocation is the resource allocation in terms of VRB pairs. The key to distributed transmission then lies in the mapping from VRB pairs to *Physical Resource Block* (PRB) pairs—that is, to the actual physical resource used for transmission.

Copyright Elsevier 2024

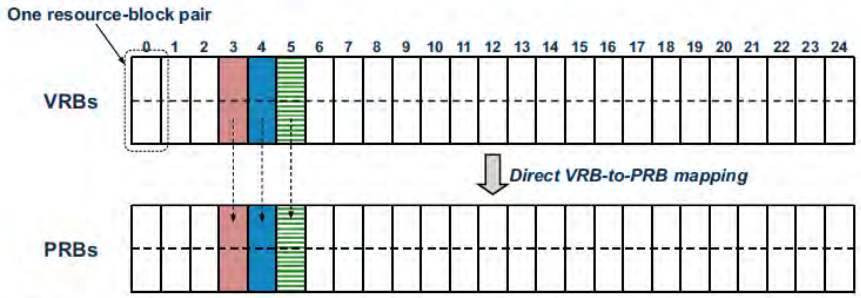


FIGURE 10.6

VRB-to-PRB mapping in case of localized VRBs. Figure assumes a cell bandwidth corresponding to 25 resource blocks

The LTE specification defines two types of VRBs: *localized* VRBs and *distributed* VRBs. In the case of localized VRBs, there is a direct mapping from VRB pairs to PRB pairs as illustrated in Figure 10.6.

However, in the case of distributed VRBs, the mapping from VRB pairs to PRB pairs is more elaborate in the sense that:

- consecutive VRBs are not mapped to PRBs that are consecutive in the frequency domain; and
- even a single VRB pair is distributed in the frequency domain.

The basic principle of distributed transmission is outlined in Figure 10.7 and consists of two steps:

- A mapping from VRB pairs to PRB pairs such that consecutive VRB pairs are not mapped to frequency-consecutive PRB pairs (first step of Figure 10.7). This provides frequency diversity between consecutive VRB pairs. The spreading in the frequency domain is done by means of a block-based "interleaver" operating on resource-block pairs.
- A split of each resource-block pair such that the two resource blocks of the resource-block pair are transmitted with a certain frequency gap in between (second step of Figure 10.7). This also provides frequency diversity for a single VRB pair. This step can be seen as the introduction of frequency hopping on a slot basis.

Whether the VRBs are localized (and thus mapped according to Figure 10.6), or distributed (mapped according to Figure 10.7) is indicated as part of the scheduling assignment in the case of type 2 resource allocation. Thus, it is possible to dynamically switch between distributed and localized transmission and also mix distributed and localized transmission for different terminals within the same subframe.

Copyright Elsevier 2024

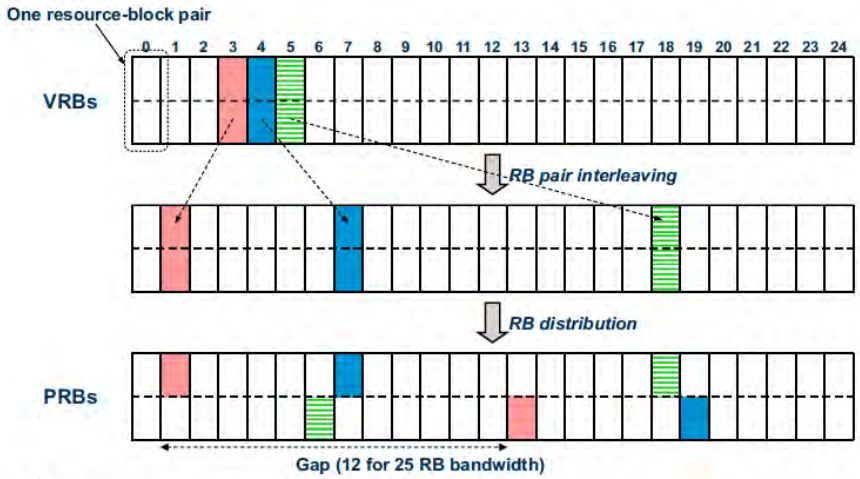


FIGURE 10.7

VRB-to-PRB mapping in case of distributed VRBs. Figure assumes a cell bandwidth corresponding to 25 resource blocks

The exact size of the frequency gap in Figure 10.7 depends on the overall downlink cell bandwidth according to Table 10.1. These gaps have been chosen based on two criteria:

- The gap should be of the order of half the downlink cell bandwidth in order to provide good frequency diversity also in the case of a single VRB pair.
- The gap should be a multiple of P^2, where P is the size of a *resource-block group* as defined in Section 10.4.6 and used for resource allocation types 0 and 1. The reason for this constraint is to ensure a smooth coexistence in the same subframe between distributed transmission as described above and transmissions based on downlink allocation types 0 and 1.

Due to the constraint that the gap size should be a multiple of P^2, the gap size will in most cases deviate from exactly half the cell bandwidth. In these cases, not all resource blocks within the cell bandwidth can be used for distributed transmission. As an example, for a cell bandwidth corresponding to 25 resource blocks (the example in Figure 10.7) and a corresponding gap size equal to 12 according to Table 10.1, the 25th resource-block pair cannot be used for distributed transmission. As another example, for a cell bandwidth corresponding to 50 resource blocks (gap size equal to 27 according to Table 10.1) only 46 resource blocks would be available for distributed transmission.

In addition to the gap size outlined in Table 10.1, for wider cell bandwidths (50 RBs and beyond), there is a possibility to use a second, smaller frequency gap with a size of the order of one-fourth of the cell bandwidth (see Table 10.2). The use of the smaller gap enables restriction of the distributed transmission to only a part of the overall cell bandwidth.

Copyright Elsevier 2024

Table 10.1 Gap Size for Different Cell Bandwidths (Number of Resource Blocks)

	Bandwidth									
	6	**7–8**	**9–10**	**11**	**12–19**	**20–26**	**27–44**	**45–63**	**64–79**	**80–110**
P	1	1	1	2	2	2	3	3	4	4
Gap size	3	4	5	4	8	12	18	27	32	48

Table 10.2 Second Gap Size for Different Cell Bandwidths (Only Applicable for Cell Bandwidths of 50 RBs and Beyond)

	Bandwidth	
	50–63	**64–110**
Gap size	9	16

Selection between the larger gap according to Table 10.1 and the smaller gap according to Table 10.2 is indicated by an additional bit in the resource allocation.

10.2 Downlink reference signals

Downlink reference signals are predefined signals occupying specific resource elements within the downlink time–frequency grid. The LTE specification includes several types of downlink reference signals transmitted in different ways and intended to be used for different purposes by a receiving terminal:

- *Cell-specific reference signals* (CRS) are transmitted in every downlink subframe and in every resource block in the frequency domain. CRS are intended to be used by terminals for channel estimation for coherent demodulation of all downlink physical channels except PMCH, PDSCH in case of *transmission modes* 7 to 10, and the EPDCCH control channel introduced in LTE release 11 (see Section 10.4).[5] CRS are also assumed to be used to acquire *channel-state information* (CSI) by terminals configured in transmission modes 1 to 8. Finally, terminal measurements on cell-specific reference signals are assumed to be used as the basis for cell-selection and handover decisions.
- *Demodulation reference signals* (DM-RS), also sometimes referred to as *UE-specific reference signals*, are intended to be used by terminals for channel estimation for coherent demodulation of PDSCH in case of transmission modes 7 to 10.[6] DM-RS are also to be used for demodulation of the EPDCCH physical channel. The alternative label

[5]See Section 10.3.1 for more details on LTE *transmission modes*.
[6]In the LTE specifications, these reference signals are actually referred to as *UE-specific reference signals,* although they are still "abbreviated" DM-RS.

Copyright Elsevier 2024

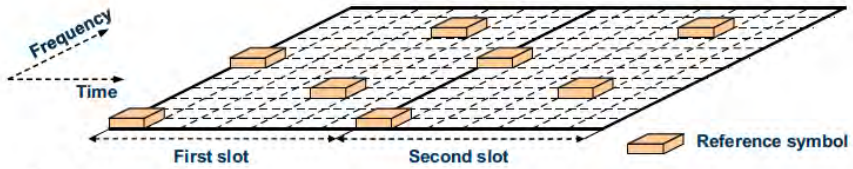


FIGURE 10.8

Structure of cell-specific reference signal within a pair of resource blocks

"*UE-specific reference signals*" relates to the fact that a specific demodulation reference signal is typically intended to be used for channel estimation by a specific terminal (=UE). The reference signal is then only transmitted within the resource blocks specifically assigned for PDSCH/EPDCCH transmission to that terminal.

- *CSI reference signals* (CSI-RS) are intended to be used by terminals to acquire channel-state information (CSI). More specifically, CSI-RS are intended to be used to acquire CSI by terminals configured in transmission modes 9 and 10. CSI-RS have a significantly lower time/frequency density, thus implying less overhead, and a higher degree of flexibility compared to the cell-specific reference signals.
- *MBSFN reference signals* are intended to be used by terminals for channel estimation for coherent demodulation in case of MCH transmission using *MBSFN* (see Chapter 17 for more details on MCH transmission).
- *Positioning reference signals* were introduced in LTE release 9 to enhance *LTE positioning functionality*, more specifically to support the use of terminal measurements on multiple LTE cells to estimate the geographical position of the terminal. The positioning reference symbols of a certain cell can be configured to correspond to empty resource elements in neighboring cells, thus enabling high-SIR conditions when receiving neighbor-cell positioning reference signals.

10.2.1 Cell-specific reference signals

Cell-specific reference signals (CRS), introduced in the first release of LTE (release 8), are the most basic downlink reference signals in LTE. There can be one, two, or four cell-specific reference signals in a cell, defining one, two, or four corresponding antenna ports, referred to as antenna port 0 to antenna port 3 in the LTE specifications.

10.2.1.1 *Structure of a single reference signal*

Figure 10.8 illustrates the structure of a single cell-specific reference signal. As can be seen, it consists of *reference symbols* of predefined values inserted within the first and third last[7]

[7]This corresponds to the fifth and fourth OFDM symbols of the slot for normal and extended cyclic prefixes respectively.

Copyright Elsevier 2024

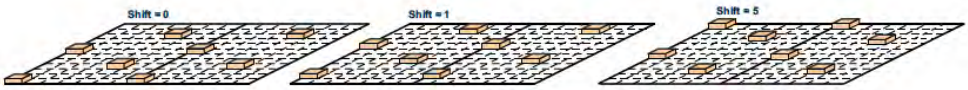


FIGURE 10.9

Different CRS frequency shifts

OFDM symbol of each slot and with a frequency-domain spacing of six subcarriers. Furthermore, there is a frequency-domain staggering of three subcarriers for the reference symbols within the third last OFDM symbol. Within each resource-block pair, consisting of 12 subcarriers during one 1 ms subframe, there are thus eight reference symbols.

In general, the values of the reference symbols vary between different reference-symbol positions and also between different cells. Thus, a cell-specific reference signal can be seen as a two-dimensional cell-specific sequence. The period of this sequence equals one 10 ms frame. Furthermore, regardless of the cell bandwidth, the reference-signal sequence is defined assuming the maximum possible LTE carrier bandwidth corresponding to 110 resource blocks in the frequency domain. Thus, the basic reference-signal sequence has a length of 8800 symbols.[8] For cell bandwidths less than the maximum possible value, only the reference symbols within that bandwidth are actually transmitted. The reference symbols in the center part of the band will therefore be the same, regardless of the actual cell bandwidth. This allows for the terminal to estimate the channel corresponding to the center part of the carrier, where, for example, the basic system information of the cell is transmitted on the BCH transport channel, without knowing the cell bandwidth. Information about the actual cell bandwidth, measured as number of resource blocks, is then provided on the BCH.

There are 504 different reference-signal sequences defined for LTE, where each sequence corresponds to one of 504 different *physical-layer cell identities*. As will be described in more detail in Chapter 14, during the so-called *cell-search procedure* the terminal detects the physical-layer identity of the cell as well as the cell frame timing. Thus, from the cell-search procedure, the terminal knows the reference-signal sequence of the cell (given by the physical-layer cell identity) as well as the start of the reference-signal sequence (given by the frame timing).

The set of reference-symbol positions outlined in Figure 10.8 is only one of six possible *frequency shifts* for the CRS reference symbols, as illustrated in Figure 10.9. The frequency shift to use in a cell depends on the physical-layer identity of the cell such that each shift corresponds to 84 different cell identities. Thus, the six different frequency shifts jointly cover all 504 different cell identities. By properly assigning physical-layer cell identities to different cells, different reference-signal frequency shifts may be used in neighboring cells. This can be beneficial, for example, if the reference symbols are transmitted with higher energy compared to other resource elements, also referred to as *reference-signal power boosting*, in order to improve the reference-signal SIR. If reference signals of neighboring

[8]Four reference symbols per resource block, 110 resource blocks per slot, and 20 slots per frame.

Copyright Elsevier 2024

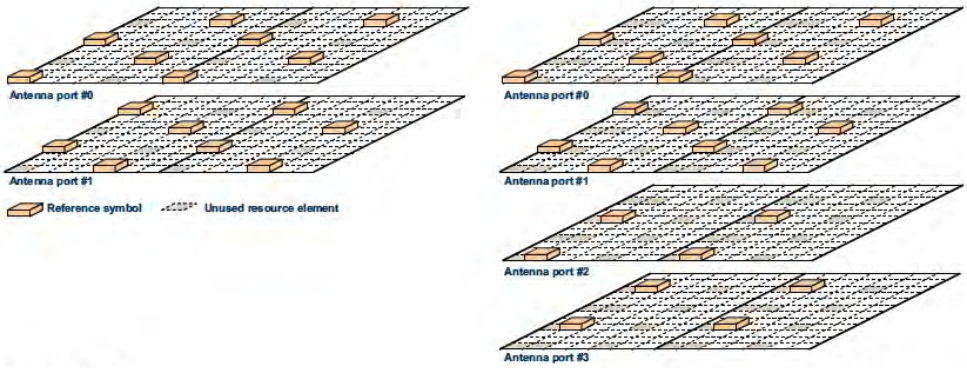


FIGURE 10.10

Structure of cell-specific reference signals in case of multiple reference signals: two reference signals corresponding to two antenna ports (left) and four reference signals corresponding to four antenna ports (right)

cells were transmitted using the same time/frequency resource, the boosted reference symbols of one cell would be interfered by equally boosted reference symbols of all neighboring cells,[9] implying no gain in the reference-signal SIR. However, if different frequency shifts are used for the reference-signal transmissions of neighboring cells, the reference symbols of one cell will at least partly be interfered by non-reference symbols of neighboring cells, implying an improved reference-signal SIR in the case of reference-signal boosting.

10.2.1.2 *Multiple reference signals*

Figure 10.10 illustrates the reference-signal structure in the case of multiple, more specifically two and four, cell-specific reference signals, and corresponding multiple antenna ports, within a cell:[10]

- In the case of two reference signals within a cell (left part of Figure 10.10), the second reference signal is frequency multiplexed with the first reference signal, with a frequency-domain offset of three subcarriers.
- In the case of four reference signals (right part of Figure 10.10), the third and fourth reference signals are frequency multiplexed and transmitted within the *second* OFDM symbol of each slot, thus being time multiplexed with the first and second reference signals.

Obviously, the reference-symbol density for the third and fourth reference signals is lower, compared to the density of the first and second reference signals. The reason for this is to

[9]This assumes that the cell transmissions are frame-timing aligned.
[10]It is not possible to configure a cell with three cell-specific reference signals.

Copyright Elsevier 2024

reduce the reference-signal overhead in the case of four reference signals. More specifically, while the first and second reference signals each correspond to a relative overhead of approximately 5% (four reference symbols within a resource block consisting of a total of 84 resource elements), the relative overhead of the third and fourth reference signals is only half of that or approximately 2.5%. This obviously has an impact on the possibility for the terminal to track very fast channel variations. However, this can be justified based on an expectation that, for example, high-order spatial multiplexing will mainly be applied to scenarios with low mobility.

It can also be noted that in a resource element carrying reference signals for a certain transmission port, nothing is being transmitted on the antenna ports corresponding to the other reference signals. Thus, a cell-specific reference signal is not interfered by transmissions on other antenna ports. Multi-antenna transmission schemes, such as spatial multiplexing, to a large extent rely on good channel estimates to suppress interference between the different layers at the receiver side. However, in the channel estimation itself there is no such suppression. Reducing the interference to the reference signals of an antenna port is therefore important in order to allow for good channel estimation, and corresponding good interference suppression, at the receiver side.

Note that, in MBSFN subframes, only the reference signals in the two first OFDM symbols of the subframe, corresponding to the control region of the MBSFN subframe, are actually transmitted. Thus, there is no transmission of cell-specific reference signals within the MBSFN part of the MBSFN subframe.

10.2.2 Demodulation reference signals

In contrast to cell-specific reference signals, a demodulation reference signal (DM-RS) is intended to be used for channel estimation by a specific terminal and is then only transmitted within the resource blocks assigned for transmission to that terminal.

DM-RS was supported in the first release of LTE (release 8). However, the use of DM-RS was then limited to the demodulation of single-layer PDSCH transmission—that is, no spatial multiplexing—corresponding to transmission mode 7. In LTE release 9, transmission based on DM-RS was extended to support dual-layer PDSCH transmission corresponding to transmission mode 8, requiring up to two simultaneous reference signals (one for each layer). Transmission based on DM-RS was then further extended in LTE release 10 to support up to eight-layer PDSCH transmission (transmission mode 9 and, from release 11, also transmission mode 10), corresponding to up to eight reference signals.[11]

Actually, the dual-layer-supporting DM-RS structure introduced in LTE release 9 was not a straightforward extension of the release-8 single-layer-limited DM-RS structure, but rather a new structure, supporting both single-layer and dual-layer transmission. Already at the time

[11]Transmission mode 10 is a release-11 extension of transmission mode 9 introducing improved support for multi-point coordination/transmission (CoMP). See also Section 10.3.1.

Copyright Elsevier 2024

Samsung Exhibit 1017, Page 201 of 538
Samsung Electronics Co., Ltd. v. Asus Technology Licensing Inc.
IPR2024-01217

of finalizing LTE release 9 it was relatively clear that the LTE radio-access technology should be further extended to support up to eight-layer spatial multiplexing in release 10. It was also quite clear that this extension would be difficult to achieve based on the release-8DM-RS structure. Rather than extending the release-8 structure to support two reference signals and then introduce a completely new structure for release 10, it was instead decided to introduce a new, more future-proof structure already in release 9. Here we will focus on the DM-RS structure introduced in LTE release 9 including the release-10 extension to support up to eight simultaneous reference signals.

The structure of the DM-RS for EPDCCH is very similar to that of DM-RS for PDSCH, although with some limitations such as support for a maximum of four reference signals.

10.2.2.1 DM-RS for PDSCH

Figure 10.11 illustrates the DM-RS time/frequency structure for the case of one or two reference signals.[12] As can be seen, there are 12 reference symbols within a resource-block pair. In contrast to cell-specific reference signals, for which the reference symbols of one reference signal correspond to unused resource elements for other reference signals (see Figure 10.10), in the case of two demodulation reference signals all 12 reference symbols in Figure 10.11 are transmitted for both reference signals. Interference between the reference signals is instead avoided by applying mutually orthogonal patterns, referred to as *Orthogonal Cover Codes* (OCC), to pairs of consecutive reference symbols as illustrated in the lower-right part of the figure.

Figure 10.12 illustrates the extended DM-RS structure introduced in LTE release 10 to support up to eight reference signals. In this case, there are up 24 reference-symbol positions within a resource-block pair. The reference signals are frequency multiplexed in groups of up to four reference signals while, within each group, the up to four reference signals are separated by means of orthogonal cover codes spanning four reference symbols in the time domain (two pairs of consecutive reference symbols). It should be noted that orthogonality between the full set of eight reference signals requires that the channel does not vary over the four reference symbols spanned by the cover codes. As the four reference symbols that the cover codes span are not consecutive in time, this implies somewhat stronger constraints in terms of the amount of channel variations that can be tolerated without seriously impacting the reference-signal orthogonality. However, more than four demodulation reference signals is only transmitted in case of spatial multiplexing with more than four layers, which is typically only applicable to low-mobility scenarios. Also note that the cover codes are defined such that, for four or less reference signals, orthogonality is achieved already over pairs of reference symbols. Thus, for three and four reference signals, the constraints on channel variations are the same as for two reference signals (Figure 10.11).

[12]In the case of TDD, the DM-RS structure is slightly modified in the DwPTS due to the shorter duration of the DwPTS compared with normal downlink subframes.

Copyright Elsevier 2024

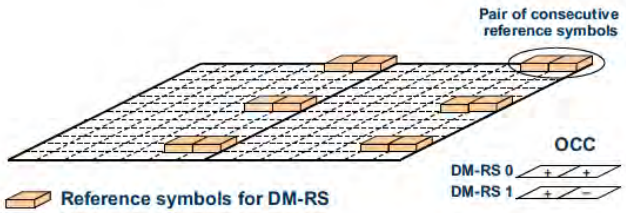


FIGURE 10.11

Structure of demodulation reference signals for the case of one or two reference signals including size-two orthogonal cover codes to separate the two reference signals

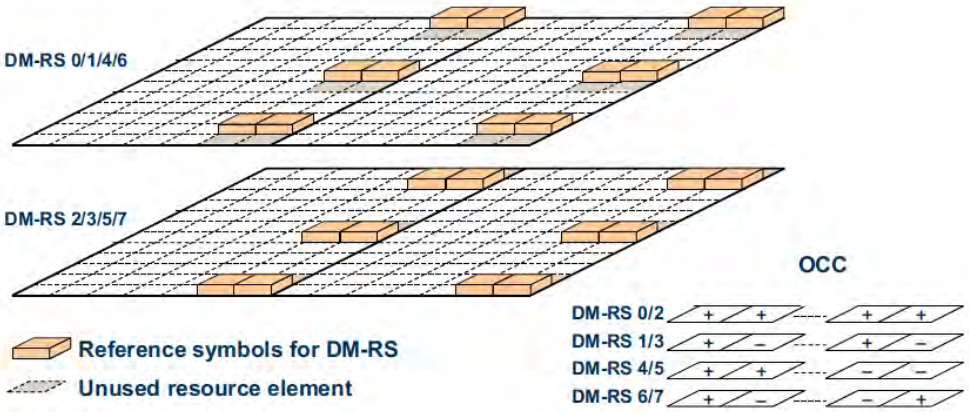


FIGURE 10.12

Demodulation reference signal structure for the case of more than two reference signals, including size-four orthogonal cover codes to separate up to four reference signals

The up to eight different PDSCH DM-RS that can be configured for a terminal correspond to antenna port 7 up to antenna port 14 in the LTE specification, with antenna port 7 and antenna port 8 corresponding to the release 9 DM-RS supporting up to two layers of spatial multiplexing.[13]

The DM-RS *reference-signal sequence*—that is, the values to be taken by the DM-RS reference symbols—repeats itself every frame. Up to and including LTE release 10, the reference-signal sequence was independent of what terminal the DM-RS transmission was intended for but depended on the physical-layer cell identity. The reference-signal sequence thus differed between cells. Furthermore, there was the possibility to dynamically—on a subframe basis—select between two different reference-signal sequences. Information about

[13]The single DM-RS supported in release 8 corresponds to antenna port 5.

Copyright Elsevier 2024

the selected sequence was then signaled to the terminal by means of a one-bit indicator in the scheduling assignment (see also Section 10.4.6). The reason for having the possibility to dynamically select between two reference-signal sequences was to be able to use the same resource block for PDSCH transmission to two different terminals and rely on spatial separation, for example, by means of beam-forming, to separate the two transmissions. Such spatial separation, in 3GPP referred to as *Multi-User MIMO* (MU-MIMO), is typically not perfect in the sense that there will remain some interference between the transmissions. By applying different reference-signal sequences for the two spatially separated transmissions, interference randomization is achieved for the channel estimation. Downlink MU-MIMO is discussed in more detail in Section 10.3, in which the different means to separate the demodulation reference signals of different transmissions are also discussed.

However, in LTE release 11, the possibility for *terminal-specific* reference-signal sequences was introduced. This was done by introducing the possibility for the network to explicitly configure a terminal with a parameter that, if configured, should replace that cell identity when deriving the reference-signal sequence to be used by the terminal. If no terminal-specific parameter is configured, the terminal should assume cell-specific reference-signal sequences in line with releases 9/10 as previously discussed.[14]

The reason for introducing the possibility for terminal-specific reference-signal sequences was to be able to spatially separate significantly more terminals within a cell. Especially in the case of so-called *shared-cell* heterogeneous deployments there may be situations with a large number of low-power transmission points, all being part of the same cell. In such a case one typically wants to be able to reuse the same physical resources—that is, the same resource blocks, for simultaneous PDSCH transmission to different terminals from several of the transmission points. For robust channel estimation, the reference signals of each of these transmissions should preferably be based on unique reference-signal sequences, thus making terminal-specific reference signal sequences desirable. Heterogeneous deployments for LTE will be extensively discussed in Chapter 16.

When DM-RS are transmitted within a resource block, PDSCH mapping to the time–frequency grid of the resource block will be modified to avoid the resource elements in which the reference signals are transmitted (the 12 and 24 resource elements in Figure 10.11 and Figure 10.12 respectively). Although this modified mapping is not "understood" by earlier-release terminals not supporting DM-RS, this is not a problem as DM-RS will only be transmitted in resource blocks that are scheduled for PDSCH transmission to terminals of later releases supporting DM-RS and thus "understanding" the modified PDSCH mapping.

As the number of transmitted layers may vary dynamically, the number of transmitted DM-RS may also vary. Thus, the transmission may dynamically change between the DM-RS structures outlined in Figures 10.11 and 10.12 respectively. The terminal is informed about the

[14] *Cell-specific reference-signal sequences for DM-RS* should not be mixed up with *cell-specific reference signals* (CRS).

Copyright Elsevier 2024

number of transmitted layers (the "transmission rank") as part of the scheduling assignment and will thus know the DM-RS structure and associated PDSCH mapping for each subframe.

10.2.2.2 DM-RS for EPDCCH

As mentioned earlier, as part of LTE release 11, a new L1/L2 control channel structure was introduced based on the so-called *enhanced PDCCH* (EPDCCH). The EPDCCH will be extensively described in Section 10.4.4. Here it can just be said that, in contrast to the legacy control-channel structure (PDCCH), the EPDCCH is transmitted within resource blocks in a similar way as PDSCH. Furthermore, in contrast to the PDCCH, EPDCCH demodulation is assumed to be based on DM-RS transmitted together with the EPDCCH, similar to the use of DM-RS for PDSCH.

The structure of DM-RS for EPDCCH is very similar to the PDSCH DM-RS structure described previously. Especially, the time/frequency structure of the EPDCCH DM-RS is the same as that for PDSCH. However, for EPDCCH, there can only be up to four DM-RS, compared to up to eight DM-RS in case of PDSCH transmission. Thus the four orthogonal covers corresponding to DM-RS 4-7 in Figure 10.11 are not supported for EPDCCH transmission. Furthermore, the EPDCCH reference-signal sequence is *always* terminal-specific—that is, the terminal is explicitly configured with a parameter that is used to derive the reference-signal sequence. It should be noted that this configuration of the DM-RS reference-signal sequence for EPDCCH is done independently of the corresponding configuration for the PDSCH DM-RS.

The antenna ports corresponding to up to four DM-RS for EPDCCH are, in the LTE specifications, referred to as antenna port 107 to antenna port 110. It should be noted that, although up to four different demodulation reference signals and corresponding antenna ports can be defined for EPDCCH, a specific EPDCCH is only transmitted from a single antenna port in case of localized transmission and two antenna ports in case of distributed transmission (see Section 10.4.4). Thus, in some sense it is somewhat misleading for the specification to talk about *up to four antenna ports* for EPDCCH, as a terminal will only see one or two DM-RS-related antenna port(s).

10.2.3 CSI reference signals

*CSI reference signal*s (CSI-RS) were introduced in LTE release 10. CSI-RS are specifically intended to be used by terminals to acquire channel-state information (CSI), for example for channel-dependent scheduling, link adaptation, and transmission settings related to multi-antenna transmission. More specifically, CSI-RS are to be used to acquire CSI for terminals configured with transmission mode 9 and 10.[15]

[15]The reason why CSI-RS are not used for transmission modes 7 and 8, despite the fact that these transmission modes also assume DM-RS for channel estimation, was simply that these transmission modes were introduced in LTE releases 8 and 9 respectively while CSI-RS was not introduced until LTE release 10.

Copyright Elsevier 2024

As already mentioned early in Section 10.2, the cell-specific reference signals (CRS), available since the first release of LTE, can also be used to acquire channel-state information. The direct reason to introduce CSI-RS was the introduction of support for up to eight-layers spatial multiplexing in LTE release 10 and the corresponding need for terminals to be able to acquire CSI for at least up to eight antenna ports.

But there was also a more fundamental desire to separate two different functions of downlink reference-signals:

- The function to acquire detailed channel estimates for coherent demodulation of different downlink transmissions.
- The function to acquire channel-state information for, for example, downlink link adaptation and scheduling.

For the early releases of LTE, both these functions relied on CRS. As a consequence, CRS has to be transmitted with high-density in both time and frequency to support accurate channel estimation and coherent demodulation also for rapidly varying channels. At the same time, in order to allow for terminals to acquire channel-state information at regular intervals, CRS has to be transmitted in every subframe regardless of whether or not there is any data transmission. For the same reason CRS is transmitted over the entire cell area and cannot be beam-formed in the direction of a specific terminal.

By introducing separate sets of reference signals for channel estimation and for the acquisition of channel-state information (DM-RS and CSI-RS respectively), more opportunities for optimization and a higher degree of flexibility are achieved. The high-density DM-RS are only transmitted when there is data to transmit and can, for example, be subject to more or less arbitrary beam-forming. At the same time, CSI-RS provides a very efficient tool for deriving channel-state information for a more arbitrary number of network nodes and antenna ports. This is especially important for the support for multi-point coordination/transmission and heterogeneous deployments, as will be further discussed in Chapters 15 and 16 respectively.

10.2.3.1 CSI-RS structure

The structure of the CSI-RS to be used by a terminal is given by a *CSI-RS configuration* defining up to eight different CSI-RS for a terminal. The possibility for up to eight CSI-RS is directly related to the support for up to eight layers spatial multiplexing and corresponding up to eight demodulation reference signals introduced in release 10. It should be noted though that the antenna ports corresponding to CSI-RS are not the same as the antenna ports corresponding to DM-RS. Antenna ports corresponding to CSI-RS typically correspond to actual transmit antennas while antenna ports corresponding to DM-RS may include any antenna precoding applied at the transmitter side (see also Section 10.3.4). The antenna ports corresponding to CSI-RS are referred to as antenna port 15 up to antenna port 22 in the LTE specification.

In the time domain, CSI-RS can be configured for transmission with different periodicity, ranging from a period of 5 ms (twice every frame) to 80 ms (every eighth frame). Furthermore, for a given CSI-RS periodicity, the exact subframe in which CSI-RS is transmitted can

Copyright Elsevier 2024

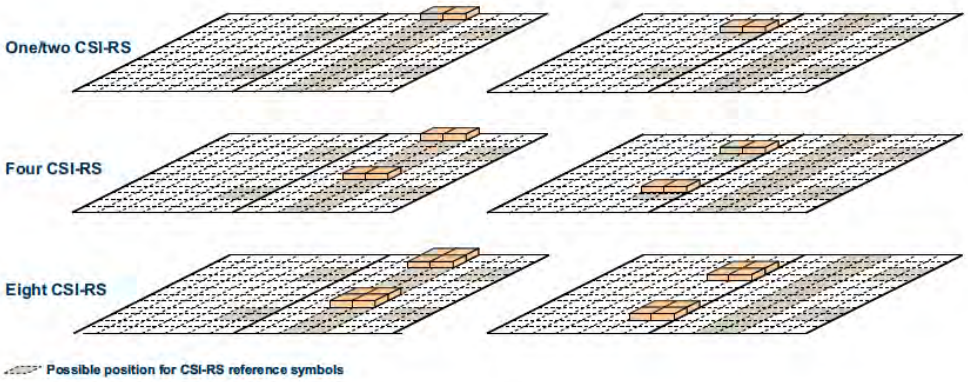


FIGURE 10.13

Examples of reference-signal positions for different number of CSI-RS within a cell. In case of a single CSI-RS, the same structure as for two CSI-RS is used

also be configured by means of a *subframe offset*.[16] In subframes in which CSI-RS is to be transmitted, it is transmitted in every resource block in the frequency domain. In other words, a CSI-RS transmission covers the entire cell bandwidth.

Within a resource-block pair, different resource elements can be used for CSI-RS transmission (illustrated by the 40 different resource elements colored gray in Figure 10.13; for TDD there are even more possibilities). Exactly what set of resource elements is used for a certain CSI-RS then depends on the exact CSI-RS configuration. More specifically:

- In the case of a CSI-RS configuration consisting of one or two configured CSI-RS, a CSI-RS consists of two consecutive reference symbols, as illustrated in the upper part of Figure 10.13. In the case of two CSI-RS, the CS-RS are then separated by applying size-two orthogonal cover codes to the two reference symbols, similar to demodulation reference signals. Thus, for the case of one or two CSI-RS, there is a possibility for 20 different CSI-RS configurations in a resource-block pair, two of which are illustrated in Figure 10.13.
- In the case of a CSI-RS configuration consisting of four/eight configured CSI-RS, the CSI-RS are pair-wise frequency multiplexed, as illustrated in the middle/lower part of Figure 10.13. For four/eight CSI-RS there is thus the possibility for ten/five different CSI-RS configurations.

To summarize, a CSI-RS configuration specifies:

- The number of CSI-RS (one, two, four, or eight).
- The CSI-RS periodicity (5 ms, 10 ms, 20 ms, 40 ms, or 80 ms).

[16] All up to eight CSI-RS of a CSI-RS configuration are transmitted within the same set of subframes—that is, with the same period and subframe offset.

Copyright Elsevier 2024

- The CSI-RS subframe offset within the CSI-RS period.
- The exact CSI-RS configuration within a resource-block pair—that is, exactly what resource elements from the 40 possible resource elements (grey resource elements in Figure 10.13) are used for the up to eight CSI-RS in a resource-block pair. There are 20, 10, and 5 different configurations for one/two, four, and eight CSI-RS respectively.

CSI-RS configurations are terminal-specific—meaning that each terminal is separately provided with a specific CSI-RS configuration that defines the number of CSI-RS to be used by the terminal and their detailed structure, as described previously. Note though that this does not mean that a certain transmitted CSI-RS is only used by one single terminal. Even if each terminal is separately provided with its CSI-RS configuration, the configuration will, in practice, be identical for a group of, or even all, terminals within a cell, implying that the terminals will, in practice, use the same set of CSI-RS to acquire CSI. However, the separate configuration of CSI-RS for different terminals *allows* for terminals within a cell to use different CSI-RS. This is important, for example, in case of shared-cell heterogeneous deployments (see Chapter 16).

10.2.3.2 CSI-RS and PDSCH mapping

As mentioned in Section 10.2.2.1, when demodulation reference signals are transmitted within a resource block, the corresponding resource elements on which the reference signals are transmitted are explicitly avoided when mapping PDSCH symbols to resource elements. This "modified" PDSCH mapping, which is not "understood" by earlier-release terminals, is possible as demodulation reference signals can be assumed to be transmitted only in resource blocks in which terminals supporting such reference signals are scheduled—that is, terminals based on LTE 10 or later.[17] Expressed alternatively, an earlier-release terminal can be assumed never to be scheduled in a resource block in which demodulation reference signals are transmitted and thus in which the modified PDSCH mapping is used.

The situation is different for CSI-RS. As a CSI-RS is transmitted within all resource blocks in the frequency domain, it would imply a strong scheduler constraint to assume that release-8/9 terminals would never be scheduled in a resource block in which CSI-RS is transmitted. If the PDSCH mapping were modified to explicitly avoid the resource elements in which CSI-RS is transmitted, the mapping would not be recognized by a releases-8/9 terminal. Instead, in the case of resource blocks scheduled to release-8/9 terminals, the PDSCH is mapped exactly according to release 8—that is, the mapping is not modified to avoid the resource elements on which CSI-RS is to be transmitted. The CSI-RS is then simply transmitted on top of the corresponding PDSCH symbols.[18] This will impact the PDSCH demodulation performance, as some PDSCH symbols will be highly corrupted. However, the

[17]Partly also for terminals of release 9, but then only for a maximum of two DM-RS.

[18]In practice, the base station may instead not transmit PDSCH at all or, equivalently, transmit PDSCH with zero energy in these resource elements in order to avoid interference to the CSI-RS transmission. The key thing is that the mapping of the remaining PDSCH symbols is in line with release 8.

Copyright Elsevier 2024

remaining PDSCH symbols will not be impacted and the PDSCH will still be decodable, although with somewhat reduced performance.

On the other hand, if a release-10 terminal is scheduled in a resource block in which CSI-RS is transmitted, the PDSCH mapping is modified to explicitly avoid the resource elements on which the CSI-RS is transmitted, similar to demodulation reference signals. Thus, if CSI-RS is transmitted in a resource block, the PDSCH mapping to that resource block will be somewhat different depending on the release of the terminal being scheduled in the resource block.

It should be noted that release-8 mapping also has to be used for transmission of, for example, system information and paging messages, as such transmissions must be possible to receive by release-8/9 terminals.

10.2.3.3 Zero-power CSI-RS

As described above, release-10-and-beyond terminals can assume that the PDSCH mapping avoids the resource elements corresponding to the set of CSI-RS configured for the terminal.

In addition to conventional CSI-RS, there is also the possibility to configure a terminal with a set of *zero-power CSI-RS* resources, where each zero-power CSI-RS has the same structure as a "conventional" (non-zero-power) CSI-RS:

- A certain periodicity (5 ms, 10 ms, 20 ms, 40 ms, or 80 ms).
- A certain subframe offset within the period.
- A certain configuration within a resource-block pair.

The intention with the zero-power CSI-RS is simply to define additional CSI-RS resources to which the terminal should assume that PDSCH is not mapped. These resources may, for example, correspond to CSI-RS of other terminals within the cell or within neighbor cells. They may also correspond to so-called CSI-IM resources, as discussed in more detail in Chapter 13.

It should be noted that, despite the name, the zero-power CSI-RS resources may not necessarily be of zero power as they may, for example, correspond to "normal" (non-zero-power) CSI-RS configured for other terminals within the cell. The key thing is that a terminal for which a certain zero-power CSI-RS resource has been configured should assume that PDSCH mapping avoids the corresponding resource elements.

10.2.4 Quasi-co-location relations

Chapter 9 briefly discussed the concept of quasi-co-located antenna ports. As also mentioned, at least for the downlink, an antenna port can be seen as corresponding to a specific reference signal. Thus it is important to understand what assumptions can be made regarding the relations, in terms of quasi-co-location, between downlink antenna ports corresponding to different reference signals.

Downlink antenna ports 0 to 3, corresponding to up to four cell-specific reference signals (CRS), can always be assumed to be jointly quasi-co-located. Similarly, antenna ports 7 to 14, corresponding to up to eight demodulation reference signals (DM-RS), can also always be assumed to be jointly quasi-located. It should be pointed out though that the quasi-co-location

Copyright Elsevier 2024

assumption for DM-RS is only valid within a subframe. The reason for this restriction is to be able to switch a PDSCH transmission that relies on DM-RS between different transmission points on a subframe basis, implying that quasi-co-location cannot be assumed *between* subframes even for a certain antenna port. Finally, antenna ports 15 to 22, corresponding to the up to eight CSI-RS of a specific CSI-RS configuration, can also always be assumed to be jointly quasi-co-located.

When it comes to quasi-co-location relations between antenna ports corresponding to different *types* of reference signals, for transmission modes 1–9 it can always be assumed that antenna ports 0–3 and 7–22—that is, CRS, DM-RS, and CSI-RS—are all *jointly* quasi-co-located. As a consequence, the only case when quasi-co-location can not necessarily be assumed for different types of reference signals is for the case of transmission mode 10. As will also be discussed in Section 10.3, transmission mode 10 was specifically introduced in LTE release 10 to support *multi-point* coordination/transmission. It is also in this case that the concept of quasi-co-location and lack thereof becomes relevant and, as indicated in Chapter 9, the concept of quasi-co-location was introduced in LTE release 11 for this specific reason. The specific aspects of quasi-co-location in case of transmission mode 10, and especially the quasi-co-location relation between the CSI-RS configured for a terminal and the set of DM-RS related to PDSCH transmission for that terminal, will be discussed in Chapter 15 as part of a more detailed discussion on multi-point coordination and transmission.

10.3 Multi-antenna transmission

As illustrated in Figure 10.14, multi-antenna transmission in LTE can, in general, be described as a mapping from the output of the data modulation to a set of antenna ports. The input to the antenna mapping thus consists of the modulation symbols (QPSK, 16QAM, 64QAM) corresponding to the one or two transport blocks of a TTI.

The output of the antenna mapping is a set of symbols for each antenna port. These symbols are subsequently applied to the OFDM modulator—that is, mapped to the basic OFDM time–frequency grid corresponding to that antenna port.

10.3.1 Transmission modes

The different multi-antenna transmission schemes correspond to different so-called *transmission modes*. There are currently ten different transmission modes defined for LTE. They differ in terms of the specific structure of the antenna mapping of Figure 10.14 but also in terms of what reference signals are assumed to be used for demodulation (cell-specific reference signals or demodulation reference signals respectively) and how CSI is acquired by the terminal and fed back to the network. Transmission mode 1 corresponds to single-antenna transmission while the remaining transmission modes correspond to different multi-antenna transmission schemes, including transmit diversity, beam-forming, and spatial multiplexing.

Copyright Elsevier 2024

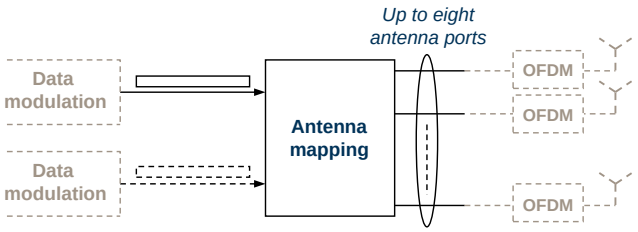


FIGURE 10.14

General structure for LTE downlink multi-antenna transmission. Modulation symbols corresponding to one or two transport blocks mapped to up to eight antenna ports

Actually, LTE supports both beam-forming and spatial multiplexing as part of more general *antenna precoding*. Furthermore, there are two approaches to downlink antenna precoding: *codebook-based precoding* and *non-codebook-based precoding*. The reason for these specific names will be further clarified below.

Transmission mode 10 is somewhat of a special case. As mentioned earlier, transmission mode 10 was introduced in LTE release 11 to support different means of dynamic multi-point coordination and transmission (see Chapter 15). From a terminal point-of-view, the downlink transmission in case of transmission mode 10 is identical to that of transmission mode 9—that is, the terminal will see an up to eight-layers PDSCH transmission and rely on DM-RS for channel estimation. One important difference between transmission mode 9 and transmission mode 10 lies in the acquisition and feedback of channel-state information (CSI) where transmission mode 10 allows for more elaborate multi-point measurements and feedback based on CSI processes, as will be further discussed in Chapter 15. Another important difference lies in what a terminal can assume in terms of quasi-co-location relations between different types of antenna ports, as mentioned in Section 10.2.4 and further discussed in Chapter 15.

It should be pointed out that transmission modes are only relevant for DL-SCH transmission. Thus, a certain transmission mode should not be seen as identical to a certain multi-antenna transmission configuration. Rather, a certain multi-antenna transmission scheme is applied to DL-SCH transmission when a terminal is configured in a certain transmission mode. The same multi-antenna transmission scheme may also be applied to other types of transmissions, such as transmission of BCH and L1/L2 control signaling.[19] However, this does not mean that the corresponding transmission mode is applied to such transmissions.

[19]Actually, only single-antenna transmission and transmit diversity is applicable to BCH and L1/L2 control signaling.

Copyright Elsevier 2024

The list below summarizes the currently defined transmission modes and the associated multi-antenna transmission schemes. The different multi-antenna transmission schemes are described in more detail in the sections below.

- **Transmission mode 1:** Single-antenna transmission.
- **Transmission mode 2:** Transmit diversity.
- **Transmission mode 3:** *Open-loop* codebook-based precoding in the case of more than one layer, transmit diversity in the case of rank-one transmission.
- **Transmission mode 4:** *Closed-loop* codebook-based precoding.
- **Transmission mode 5:** Multi-user-MIMO version of transmission mode 4.
- **Transmission mode 6:** Special case of closed-loop codebook-based precoding limited to single-layer transmission.
- **Transmission mode 7:** Non-codebook-based precoding supporting single-layer PDSCH transmission.
- **Transmission mode 8:** Non-codebook-based precoding supporting up to two layers (introduced in LTE release 9).
- **Transmission mode 9:** Non-codebook-based precoding supporting up to eight layers (extension of transmission mode 8, introduced in LTE release 10).
- **Transmission mode 10:** Extension of transmission mode 9 for enhanced support of different means of downlink multi-point coordination and transmission, also referred to as CoMP (introduced in LTE release 11).

In case of transmission modes 1 to 6, transmission is carried out from antenna ports 0 to 3. Thus the cell-specific reference signals (CRS) are to be used for channel estimation. Transmission mode 7 corresponds to transmission on antenna port 5 while transmission modes 8 to 10 correspond to transmission on antenna ports 7 to 14 (in case of transmission mode 8 limited to antenna ports 7 and 8). Thus, for transmission modes 7 to 10, demodulation reference signals (DM-RS) are to be used for channel estimation.

In practice, terminals configured for transmission modes 1 to 8 can be assumed to rely on CRS to acquire CSI while, for transmission modes 9 and 10, CSI-RS should be used.

It should also be mentioned that, although a certain multi-antenna transmission scheme can be seen as being associated with a certain transmission mode, for transmission modes 3–10 there is a possibility for fallback to transmit diversity without implying that the configured transmission mode is changed.

10.3.2 Transmit diversity

Transmit diversity can be applied to any downlink physical channel. However, it is especially applicable to transmissions that cannot be adapted to time-varying channel conditions by means of link adaptation and/or channel-dependent scheduling, and thus for which diversity is more important. This includes transmission of the BCH and PCH transport channels, as well as L1/L2 control signaling. Actually, as already mentioned, transmit diversity is the *only*

Copyright Elsevier 2024

multi-antenna transmission scheme applicable to these channels. Transmit diversity is also used for transmission of DL-SCH when transmission mode 2 is configured. Furthermore, as also already mentioned, transmit diversity is a "fallback mode" for DL-SCH transmission when the terminal is configured in transmission modes 3 and higher. More specifically, a scheduling assignment using DCI format 1A (see Section 10.4.6) implies the use of transmit diversity regardless of the configured transmission mode.

Transmit diversity assumes the use of cell-specific reference signals for channel estimation. Thus, a transmit-diversity signal is always transmitted on the same antenna ports as the cell-specific reference signals (antenna ports 0 to 3). Actually, if a cell is configured with two cell-specific reference signals, transmit diversity for two antenna ports *must be used* for BCH and PCH, as well as for the L1/L2 control signaling on PDCCH.[20] Similarly, if four cell-specific reference signals are configured for the cell, transmit diversity for four antenna ports has to be used for the transmission of these channels. In this way, a terminal does not have to be explicitly informed about what multi-antenna transmission scheme is used for these channels. Rather, this is given implicitly from the number of cell-specific reference signals configured for a cell.[21]

10.3.2.1 Transmit diversity for two antenna ports

In the case of two antenna ports, LTE transmit diversity is based on *Space-Frequency Block Coding* (SFBC), as described in Chapter 5. As can be seen from Figure 10.15, SFBC implies that two consecutive modulation symbols S_i and S_{i+1} are mapped directly to frequency-adjacent resource elements on the first antenna port. On the second antenna port the frequency-swapped and transformed symbols $-S_{i+1}^*$ and S_i^* are mapped to the corresponding resource elements, where "*" denotes complex conjugate.

Figure 10.15 also indicates how the antenna ports on which a transmit-diversity signal is being transmitted correspond to the cell-specific reference signals, more specifically CRS 0 and CRS 1 in the case of two antenna ports. Note that one should not interpret this such that the CRS is specifically transmitted for this transmit-diversity signal. There are, in practice, multiple transmissions on antenna ports 0 and 1, all of which rely on the corresponding CRS for channel estimation.

10.3.2.2 Transmit diversity for four antenna ports

In the case of four antenna ports, LTE transmit diversity is based on a combination of SFBC and *Frequency-Switched Transmit Diversity* (FSTD). As can be seen in Figure 10.16, combined SFBC/FSTD implies that pairs of modulation symbols are transmitted by means of SFBC with transmission alternating between pairs of antenna ports (antenna ports 0 and 2 and antenna ports 1 and 3 respectively). For the resource elements where transmission is on one

[20] Note that this is not true for the EPDCCH control channel.

[21] Actually, the situation is partly the opposite, i.e. the terminal blindly detects the number of antenna ports used for BCH transmission and, from that, decides on the number of cell-specific reference signals configured within the cell.

Copyright Elsevier 2024

Samsung Exhibit 1017, Page 213 of 538
Samsung Electronics Co., Ltd. v. Asus Technology Licensing Inc.
IPR2024-01217

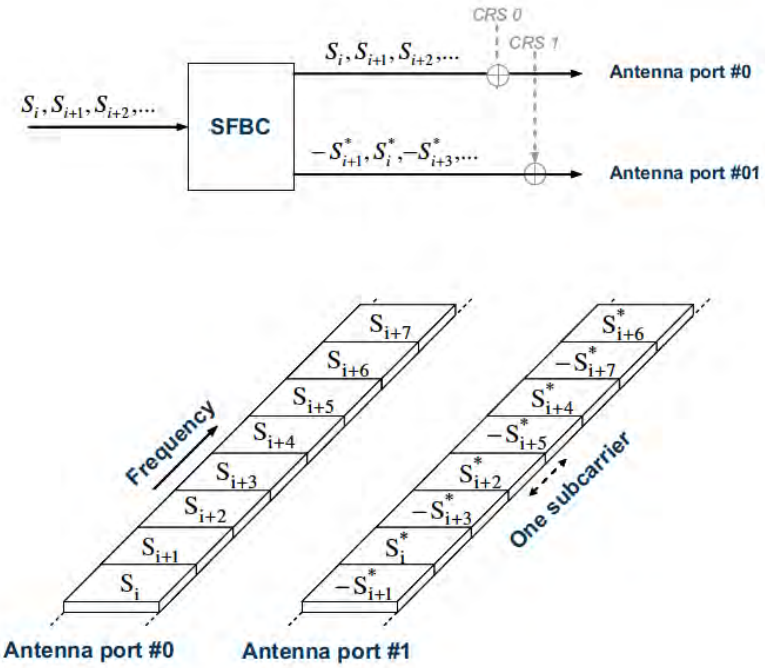


FIGURE 10.15

Transmit diversity for two antenna ports—SFBC

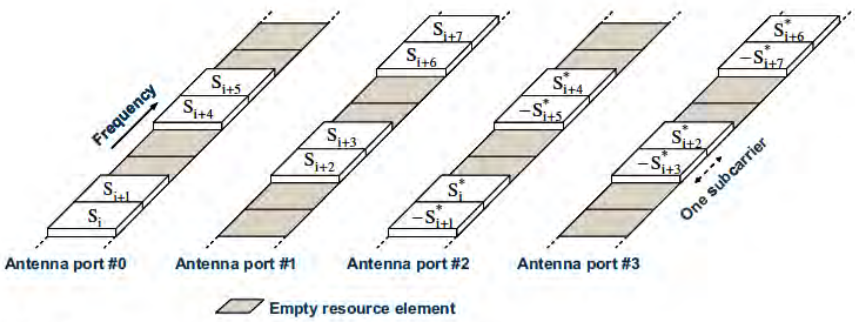


FIGURE 10.16

Transmit diversity for four antenna ports—Combined SFBC/FSTD

Copyright Elsevier 2024

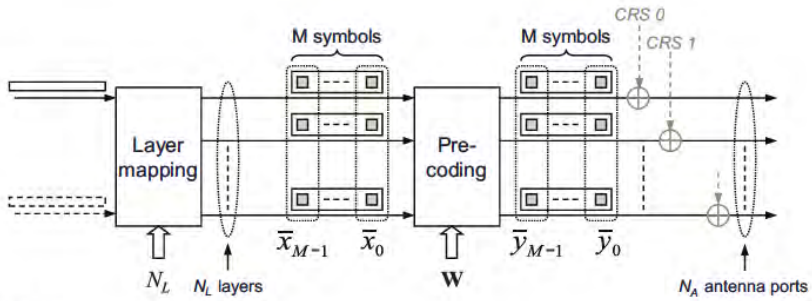


FIGURE 10.17

The basic structure of LTE codebook-based antenna precoding. The figure also indicates how cell-specific reference signals (CRS) are applied after pre-coding

pair of antenna ports, there is no transmission on the other pair of antenna ports. Thus, combined SFBC/FSTD in some sense operates on groups of four modulation symbols and corresponding groups of four frequency-consecutive resource elements on each antenna port.

10.3.3 Codebook-based precoding

The basic processing for codebook-based precoding is illustrated in Figure 10.17. The modulation symbols corresponding to one or two transport blocks are first mapped to N_L *layers*. The number of layers may range from a minimum of one layer up to a maximum number of layers equal to the number of antenna ports.[22] The layers are then mapped to the antenna ports by means of the *precoder*. As codebook-based precoding relies on the cell-specific reference signals for channel estimation, and there are at most four cell-specific reference signals in a cell, codebook-based precoding allows for a maximum of four antenna ports and, as a consequence, a maximum of four layers.

The mapping to layers is outlined in Figure 10.18 for the case of an initial transmission. There is one transport block in the case of a single layer ($N_L = 1$) and two transport blocks for two or more layers ($N_L > 1$). In the case of a hybrid-ARQ retransmission, if only one of two transport blocks needs to be retransmitted and that transport block was mapped to two layers for the initial transmission, the retransmission will also be carried out on two layers. Thus, in the case of a retransmission, a single transport block may also be transmitted using two layers.

The mapping to layers is such that the number of modulation symbols on each layer is the same and equal to the number of symbols to be transmitted on each antenna port. Thus, in the case of three layers, there should be twice as many modulation symbols corresponding to the second transport block (mapped to the second and third layers) compared to the first

[22] In practice, the number of layers is also limited by, and should not exceed, the number of receive antennas available at the terminal.

Copyright Elsevier 2024

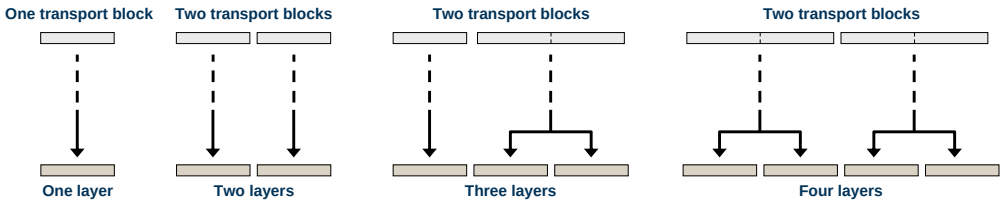


FIGURE 10.18

Transport-block-to-layer mapping for codebook-based antenna precoding (initial transmission)

transport block (mapped to the first layer). This is ensured by the selection of an appropriate transport-block size in combination with the rate-matching functionality. In the case of four layers, the first transport block is mapped to the first and second layers while the second transport block is mapped to the third and fourth layers. In this case, the number of modulation symbols should thus be the same for the two transport blocks. For one transport block being mapped to two layers, the modulated symbols corresponding to the transport block are mapped to the layers in an alternating fashion—that is, every second modulation symbol is mapped to the first and second layer respectively.

In the case of multi-antenna precoding the number of layers is also often referred to as the *transmission rank*.[23] The transmission rank can vary dynamically, for example based on the number of layers that can be supported by the channel. The latter is sometimes also referred to as the *channel rank*.

After layer mapping, a set of N_L symbols (one symbol from each layer) is linearly combined and mapped to the antenna ports. This combination/mapping can be described by a *precoder matrix* $\mathbf{W}$ of size $N_A \times N_L$, where N_A is the number of antenna ports which, for codebook-based precoding equal two or four. More specifically, the vector $\overline{y}_i$ of size N_A, consisting of one symbol for each antenna port, is given by $\overline{\mathbf{y}}_i = \mathbf{W} \cdot \overline{x}_i$, where the vector $\overline{x}_i$ of size N_L consists of one symbol from each layer. As the number of layers can vary dynamically, also the number of columns of the precoder matrix will vary dynamically. Specifically, in the case of a single layer, the precoder matrix $\mathbf{W}$ is a vector of size $N_A \times 1$ that provides beam-forming for a single modulation symbol.

Figure 10.17 also indicates how the cell-specific reference signals are applied after antenna precoding. Channel estimation based on the cell-specific reference signals will thus reflect the channel for each antenna port *not including the precoding*. As a consequence, the terminal receiver must have explicit knowledge about what precoding has been applied at the transmitter side in order to properly process the received signal and recover the different layers. Once again, the figure should not be interpreted such that cell-specific reference signals are inserted specifically for a given PDSCH transmission.

[23]In the LTE specification, transmit diversity is also described as transmission using *multiple layers*. However, transmit diversity is still a *single-rank* transmission scheme.

Copyright Elsevier 2024

There are two operational modes for codebook-based precoding, *closed-loop operation* and *open-loop operation*. These two modes differ in terms of the exact structure of the precoder matrix and how the matrix is selected by the network and made known to the terminal.

10.3.3.1 Closed-loop precoding

In case of closed-loop precoding it is assumed that the network selects the precoder matrix based on feedback from the terminal. As already mentioned, closed-loop precoding is associated with transmission mode 4.

Based on measurements on the cell-specific reference signals, the terminal selects a suitable transmission rank and corresponding precoder matrix. Information about the selected rank and precoder matrix is then reported to the network in the form of a *Rank Indication* (RI) and a *Precoder-Matrix Indication* (PMI), as described in Section 11.4. It is important to understand that the RI and PMI are only recommendations and the network does not need to follow the RI/PMI provided by the terminal when selecting the actual transmission rank and precoder matrix to be used for transmission to the terminal. When not following the terminal recommendation, the network must explicitly inform the terminal what precoder matrix is used for the downlink transmission. On the other hand, if the network uses the precoder matrix recommended by the terminal, only a confirmation that the network is using the recommended matrix is signaled.

To limit the signaling on both uplink and downlink, only a limited set of precoder matrices, also referred to as the *codebook*, is defined for each transmission rank for a given number of antenna ports. Both the terminal (when reporting PMI) and the network (when selecting the actual precoder matrix to use for the subsequent downlink transmission to the terminal) should select a precoder matrix from the corresponding codebook. Thus, for terminal PMI reporting, as well as when the network informs the terminal about the actual precoder matrix used for the downlink transmission, only the index of the selected matrix needs to be signaled.

As LTE supports codebook-based precoding for two and four antenna ports, codebooks are defined for:

- Two antenna ports and one and two layers, corresponding to precoder matrices of size 2×1 and 2×2 respectively.
- Four antenna ports and one, two, three, and four layers, corresponding to precoder matrices of size 4×1, 4×2, 4×3, and 4×4 respectively.

As an example, the precoder matrices specified for the case of two antenna ports are illustrated in Table 10.3. As can be seen, there are four 2×1 precoder matrices for single-layer transmission and three 2×2 precoder matrices for two-layer transmission. In the same way, sets of 4×1, 4×2, 4×3, and 4×4 matrices are defined for the case of four antenna ports and one, two, three, and four layers respectively. It should be pointed out that the first rank-2 (2×2) matrix in Table 10.3 is not used in closed-loop operation but only for *open-loop precoding*, as described in the next section.

Copyright Elsevier 2024

Table 10.3 Precoder Matrices for Two Antenna Ports and One and Two Layers. The First 2×2 Matrix is only Used for Open-Loop Precoding

One layer	$\frac{1}{\sqrt{2}}\begin{bmatrix}+1\\+1\end{bmatrix}$	$\frac{1}{\sqrt{2}}\begin{bmatrix}+1\\-1\end{bmatrix}$	$\frac{1}{\sqrt{2}}\begin{bmatrix}+1\\+\mathrm{j}\end{bmatrix}$	$\frac{1}{\sqrt{2}}\begin{bmatrix}+1\\-\mathrm{j}\end{bmatrix}$
Two layers	$\frac{1}{\sqrt{2}}\begin{bmatrix}+1 & 0\\0 & +1\end{bmatrix}$	$\frac{1}{2}\begin{bmatrix}+1 & +1\\+1 & -1\end{bmatrix}$	$\frac{1}{2}\begin{bmatrix}+1 & +1\\+\mathrm{j} & -\mathrm{j}\end{bmatrix}$	

Even if the network is following the precoder-matrix recommendation provided by the terminal, the network may, for different reasons, decide to use a lower rank for the transmission, so-called *rank override*. In that case the network will use a subset of the columns of the recommended precoder matrix. The network precoder confirmation will then include explicit information about the set of columns being used or, equivalently, about the set of layers being transmitted.

There is also a possibility to apply closed-loop precoding strictly limited to single-layer (rank-1) transmission. This kind of multi-antenna transmission is associated with transmission mode 6. The reason for defining an additional transmission mode limited to single-layer transmission rather than relying on the general closed-loop precoding associated with transmission mode 4 is that, by strictly limiting to single-layer transmission, the signaling overhead on both downlink and uplink can be reduced. Transmission mode 6 can, for example, be configured for terminals with low SINR for which multi-layer transmission would not apply in order to harvest the beam-forming gain.

10.3.3.2 Open-loop precoding

Open-loop precoding does not rely on any detailed precoder recommendation being reported by the terminal and does not require any explicit network signaling of the actual precoder used for the downlink transmission. Instead, the precoder matrix is selected in a predefined and deterministic way known to the terminal in advance. One use of open-loop precoding is in high-mobility scenarios where accurate feedback is difficult to achieve due to the latency in the PMI reporting. As already mentioned, open-loop precoding is associated with transmission mode 3.

The basic transmission structure for open-loop precoding is aligned with the general codebook-based precoding outlined in Figure 10.17 and only differs from closed-loop precoding in the structure of the precoding matrix $\mathbf{W}$.

In the case of open-loop precoding, the precoder matrix can be described as the product of two matrices $\mathbf{W}'$ and $\mathbf{P}$, where $\mathbf{W}'$ and $\mathbf{P}$ are of size $N_A \times N_L$ and $N_L \times N_L$ respectively:

$$\mathbf{W} = \mathbf{W}' \cdot \mathbf{P} \tag{10.1}$$

Copyright Elsevier 2024

In the case of two antenna ports, the matrix $\mathbf{W}'$ is the normalized 2×2 identity matrix:[24]

$$\mathrm{W}' = \frac{1}{\sqrt{2}} \begin{bmatrix} +1 & 0 \\ 0 & +1 \end{bmatrix} \tag{10.2}$$

In the case of four antenna ports, $\mathbf{W}'$ is given by cycling through four of the defined $4 \times N_{\mathrm{L}}$ precoder matrices and is different for consecutive resource elements.

The matrix $\mathbf{P}$ can be expressed as $\mathbf{P} = \mathbf{D}_i \cdot \mathbf{U}$, where $\mathbf{U}$ is a constant matrix of size $N_{\mathrm{L}} \times N_{\mathrm{L}}$ and $\mathbf{D}_i$ is a matrix of size $N_{\mathrm{L}} \times N_{\mathrm{L}}$ that varies between subcarriers (indicated by the index i). As an example, the matrices $\mathbf{U}$ and $\mathbf{D}_i$ for the case of two layers ($N_{\mathrm{L}} = 2$) are given by:

$$\mathbf{U} = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & e^{\,j2\pi/2} \end{bmatrix} \mathrm{D}_i = \begin{bmatrix} 1 & 0 \\ 0 & e^{\,j2\pi i/2} \end{bmatrix} \tag{10.3}$$

The basic idea with the matrix $\mathbf{P}$ is to average out any differences in the channel conditions as seen by the different layers.

Similar to closed-loop precoding, the transmission rank for open-loop precoding can also vary dynamically down to a minimum of two layers. Transmission mode 3, associated with open-loop precoding, also allows for rank-1 transmission. In that case, transmit diversity as described in Section 10.3.1 is used—that is, SFBC for two antenna ports and combined SFBC/FSTD for four antenna ports.

10.3.4 Non-codebook-based precoding

Similar to codebook-based precoding, non-codebook-based precoding is only applicable to DL-SCH transmission. Non-codebook-based precoding was introduced in LTE release 9 but was then limited to a maximum of two layers. The extension to eight layers was then introduced as part of release 10. The release-9 scheme, associated with transmission mode 8, is a subset of the extended release-10 scheme (transmission mode 9).

There is also a release-8 non-codebook-based precoding defined, associated with transmission mode 7. Transmission mode 7 relies on the release-8 demodulation reference signals mentioned but not described in detail in Section 10.2.2 and only supports single-layer transmission. In this description we will focus on non-codebook-based precoding corresponding to transmission modes 8 and 9.

The basic principles for non-codebook-based precoding can be explained based on Figure 10.19 (where the precoder is intentionally shaded). As can be seen, this figure is very similar to the corresponding figure illustrating codebook-based precoding (Figure 10.17), with layer mapping of modulation symbols corresponding to one or two transport blocks followed by precoding. The layer mapping also follows the same principles as that of codebook-based precoding (see Figure 10.18) but is extended to support up to eight layers. In particular, at least for an initial transmission, there are two transport blocks per TTI

[24] As non-codebook-based precoding is not used for rank-1 transmission, there is no need for any matrix $\mathbf{W}'$ of size 2×1.

Copyright Elsevier 2024

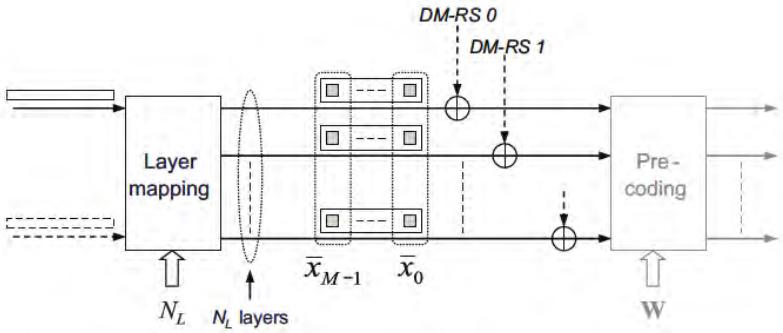


FIGURE 10.19

Basic principles for non-codebook-based antenna precoding

except for the case of a single layer, in which case there is only one transport block within the TTI. Similar to codebook-based precoding, for hybrid-ARQ retransmissions there may in some cases be a single transport block also in the case of multi-layer transmission.

The main difference in Figure 10.19 compared to Figure 10.17 (codebook-based precoding) is the presence of demodulation reference signals (DM-RS) before the precoding. The transmission of precoded reference signals allows for demodulation and recovery of the transmitted layers at the receiver side *without explicit receiver knowledge of the precoding applied at the transmitter side*. Put simply, channel estimation based on precoded demodulation reference signals will reflect the channel experienced by the layers, *including the precoding*, and can thus be used directly for coherent demodulation of the different layers. There is no need to signal any precoder-matrix information to the terminal, which only needs to know the number of layers—that is, the transmission rank. As a consequence, the network can select an arbitrary precoder and there is no need for any explicit codebook to select from. This is the reason for the term *non-codebook-based* precoding. It should be noted though that non-codebook-based precoding may still rely on codebooks for the terminal feedback, as described below.

The possibility to select an arbitrary precoder matrix for the transmission is also the reason why the precoder is shaded in Figure 10.19. The precoder part of Figure 10.19 is not visible in the LTE specification and, strictly speaking, in the case of non-codebook-based precoding the antenna mapping defined according to Figure 10.14 consists of only the layer mapping. This also means that the antenna ports defined in Figure 10.14 correspond to the different layers in Figure 10.19, or, expressed differently, precoding occurs *after* the antenna ports.

Still, there must be a way for the network to select a suitable precoder matrix for the transmission. There are essentially two ways by which this can be done in the case of non-codebook-based precoding.

Copyright Elsevier 2024

The network may estimate the uplink channel state, for example based on transmission of uplink sounding reference signals as described in the next chapter, and rely on uplink/downlink channel reciprocity when selecting the precoder matrix to use for the downlink transmission. This is especially of interest for TDD operation for which the use of the same frequency for uplink and downlink transmission typically leads to a higher degree of downlink/uplink channel reciprocity. It should be noted though that if the terminal uses multiple receive antennas, it also has to transmit on multiple antennas in order for uplink measurements to fully reflect the downlink channel state.

Alternatively, the network may rely on terminal feedback for precoder-matrix selection. For transmission modes 8 and 9, this feedback is actually very similar to the corresponding feedback for closed-loop codebook-based precoding. In particular, the feedback is based on the same precoder matrices as for codebook-based precoding (see Table 10.3 for an example) but extended to handle up to eight antenna ports. Furthermore, for transmission mode 9, the terminal measurements should be based on CSI-RS, as described in Section 10.2.3, rather than the cell-specific reference signals.

Thus, despite the name, non-codebook-based precoding may also use defined codebooks. However, in contrast to codebook-based precoding, the codebooks are then only used for the terminal PMI reporting and not for the actual downlink transmission.

10.3.5 Downlink multi-user MIMO

Spatial multiplexing implies the transmission of multiple layers—that is, multiple parallel transmissions on the same time–frequency resource—to the same terminal. The presence of multiple antennas at both the transmitter and receiver sides in combination with transmitter and/or receiver signal processing is then used to suppress interference between the different layers.

Spatial multiplexing has often been referred to as *Multi-Input-Multi-Output* (MIMO) transmission, reflecting the fact that the channel in the case of spatial multiplexing can be seen as having *multiple inputs*, corresponding to the multiple transmit antennas, and *multiple outputs*, corresponding to the multiple receive antennas. The more specific term *Single-User MIMO* (SU-MIMO) is also often used for reasons that will become clearer below.[25]

The term *Multi-User* MIMO (MU-MIMO) is, in 3GPP, used to denote transmission to *different* terminals using *the same time–frequency resource*, in practice relying on multiple antennas at the transmitter (network) side to separate the two transmissions.

In principle, one could realize MU-MIMO as a direct extension to spatial multiplexing, with the different layers simply being intended for different terminals. The set of terminals would demodulate and decode the full set of layers in the same way as for SU-MIMO.

[25]The term MIMO is used to refer to any transmission using multiple transmit antennas and multiple receive antennas—not only limited to spatial multiplexing but also, for example, single-layer precoding to get beamforming gains.

Copyright Elsevier 2024

Samsung Exhibit 1017, Page 221 of 538
Samsung Electronics Co., Ltd. v. Asus Technology Licensing Inc.
IPR2024-01217

The data on a layer not intended for a specific terminal would then just be discarded by that terminal after demodulation/decoding.

However, such an approach would imply that all terminals involved in the MU-MIMO transmission would need to know about the full set of layers being transmitted. It would also imply that one would need to have exactly the same resource assignment—that is, transmission on the same set of resource blocks, for all terminals involved in the MU-MIMO transmission. All terminals would also need to include the full set of receive antennas necessary to receive the overall multi-layer transmission.

Instead, 3GPP has chosen an MU-MIMO approach that does not require terminal knowledge about the presence of the other transmissions, allows for only partly overlapping resource assignments, and, at least in principle, does not require the presence of multiple receive antennas at the mobile terminal.[26] There are two approaches to MU-MIMO with explicit support in the LTE specifications, one being an integrated part of transmission modes 8 and 9 corresponding to non-codebook-based precoding, and one being based on codebook-based precoding but associated with a special transmission mode, *transmission mode 5*.

10.3.5.1 MU-MIMO within transmission modes 8/9

In principle, MU-MIMO based on transmission modes 8 and 9 is straightforward. Based on feedback of channel-state information from terminals within the cell, the base station selects two or more terminals to transmit to using the same set of time-frequency resources. Non-codebook-based precoding for one or in some cases even multiple layers is then applied to each transmission in such a way that they are spatially separated at the receiver (terminal) sides.

The spatial separation at the terminal side will typically not be perfect. To enhance channel estimation at the terminal side it is therefore preferred to use different demodulation reference signals (DM-RS) for the different transmissions in order to improve channel estimation.[27]

As discussed in Section 10.2.2.1, there are two methods by which reference-signal sequences can be assigned to downlink DM-RS:

- Cell-specific assignment supported from LTE release 9, i.e. the release in which transmission mode 8 and DM-RS were introduced.
- Fully terminal-specific assignment supported from LTE release 11.

As also described in Section 10.2.2.1, in case of cell-specific assignment there is the possibility to dynamically select between two different cell-specific reference-signal sequences.

[26]Note, though, that the LTE performance requirements in general assume the presence of at least two receive antennas at the mobile terminal.

[27]As discussed in Section 10.3.1, transmission modes 8 and 9 are assumed to rely on DM-RS for channel estimation.

Copyright Elsevier 2024

With pre-release-11 (cell-specific) DM-RS sequences there is the possibility for up to four different DM-RS supporting MU-MIMO between up to four different terminals.

- In case of single-layer transmission, the network can explicitly signal on which of antenna ports 7 and 8, corresponding to the two orthogonal cover codes (OCC), a transmission is carried out. This, in combination with the possibility to dynamically select between two reference-signal sequences, allows for up to four different DM-RS and a corresponding possibility for single-layer MU-MIMO transmission to up to four terminals in parallel.
- In case of dual-layer transmission (on antenna ports 7 and 8), the possibility to dynamically select between two reference-signal sequences allows for MU-MIMO transmission to up to two terminals in parallel.
- In case of more than two layers, there is no signaling support for selecting reference-signal sequence and thus no pre-release-11 support for MU-MIMO.

Note that one can also do MU-MIMO for single-layer and dual-layer transmission in parallel. There could, for example, be one two-layer transmission using one of the two reference-signal sequences and up to two one-layer transmissions using the other reference-signal sequence, separated by means of the two different OCCs.

However, with the release-11 introduction of terminal-specific assignment of reference-signal sequences, as described in Section 10.2.2.1, MU-MIMO can, at least in principle, be applied to an arbitrary number of terminals regardless of the number of layers. In a normal, for example, macro-deployment, the number of terminals for which MU-MIMO can jointly be carried out is limited by the number of transmitter-side antennas and, in practice, the possibility for MU-MIMO for up to four terminals in parallel as supported prior to release 11 is typically sufficient. However, the possibility for MU-MIMO between significantly more terminals is important in specific scenarios, especially in so-called shared-cell heterogeneous deployments (see further Chapter 16).

10.3.5.2 MU-MIMO based on CRS

The above-described MU-MIMO transmission is part of transmission modes 8 and 9 and thus became available in LTE release 9, with further extension in release 10. However, already in LTE release 8, MU-MIMO was possible by a minor modification of transmission mode 4—that is, closed-loop codebook-based beam-forming, leading to *transmission mode 5*. The only difference between transmission modes 4 and 5 is the signaling of an additional power offset between PDSCH and the cell-specific reference signals (CRS).

In general, for transmission modes relying on CRS (as well as when relying on DM-RS) for channel estimation the terminal will use the reference signal as a phase reference but also as a power/amplitude reference for the demodulation of signals transmitted by means of higher-order modulation (16QAM and 64QAM). Thus, for proper demodulation of higher-order modulation, the terminal needs to know the power offset between the CRS and the PDSCH.

Copyright Elsevier 2024

The terminal is informed about this power offset by means of higher-layer signaling. However, what is then provided is the offset between CRS power and the overall PDSCH power, including all layers. In the case of spatial multiplexing, the overall PDSCH power has to be divided between the different layers, and it is the relation between the CRS power and the per-layer PDSCH power that is relevant for demodulation.

In the case of pure spatial multiplexing (no MU-MIMO)—that is, transmission modes 3 and 4—the terminal knows about the number of layers and thus indirectly about the offset between the CRS power and the per-layer PDSCH power.

In the case of MU-MIMO, the total available power will typically also be divided between the transmissions to the different terminals, with less PDSCH power being available for each transmission. However, terminals are not aware of the presence of parallel transmissions to other terminals and are thus not aware of any per-PDSCH power reduction. For this reason, transmission mode 5 includes the explicit signaling of an *additional power offset* of −3 dB to be used by the terminal in addition to the CRS/PDSCH power offset signaled by higher layers.

Transmission mode 5 is limited to single-rank transmission and, in practice, limited to two users being scheduled in parallel, as there is only a single −3 dB offset defined.

Note that the above power-offset signaling is not needed for MU-MIMO based on demodulation reference signals (transmission modes 8 and 9) as, in this case, each transmission has its own set of reference signals. The power per reference signal will thus scale with the number of layers and transmissions, similar to the PDSCH power, and the reference-signal to per-layer PDSCH power ratio will remain constant.

10.4 Downlink L1/L2 control signaling

To support the transmission of downlink and uplink transport channels, there is a need for certain *associated downlink control signaling*. This control signaling is often referred to as *downlink L1/L2 control signaling*, indicating that the corresponding information partly originates from the physical layer (Layer 1) and partly from Layer 2 MAC. Downlink L1/L2 control signaling consists of downlink scheduling assignments, including information required for the terminal to be able to properly receive, demodulate, and decode the DL-SCH[28] on a component carrier, uplink scheduling grants informing the terminal about the resources and transport format to use for uplink (UL-SCH) transmission, and hybrid-ARQ acknowledgements in response to UL-SCH transmissions. In addition, the downlink control signaling can also be used for the transmission of power-control commands for power control of uplink physical channels, as well as for certain special purposes such as MBSFN notifications.

The basic time-frequency structure for transmission of L1/L2 control signaling is illustrated in Figure 10.20, with control signaling being located at the beginning of each subframe

[28]L1/L2 control signaling is also needed for the reception, demodulation, and decoding of the PCH transport channel.

Copyright Elsevier 2024

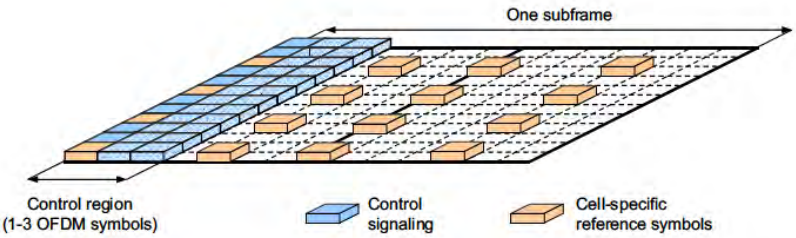


FIGURE 10.20

LTE time/frequency grid illustrating the split of the subframe into (variable-sized) control and data regions

and spanning the full downlink carrier bandwidth. Each subframe can therefore be said to be divided into a *control region* followed by a *data region*, where the control region corresponds to the part of the subframe in which the L1/L2 control signaling is transmitted. Starting from release 11, there is also a possibility to locate parts of the L1/L2 control signaling in the data region as described further below. However, the split of a subframe into a control region and a data region still applies.

To simplify the overall design, the control region always occupies an integer number of OFDM symbols, more specifically one, two, or three OFDM symbols (for narrow cell bandwidths, 10 resource blocks or less, the control region consists of two, three, or four OFDM symbols to allow for a sufficient amount of control signaling). In the case of carrier aggregation, there is one control region per component carrier.

The size of the control region expressed in number of OFDM symbols, or, equivalently, the start of the data region, can be dynamically varied on a per-subframe basis independently for each component carrier.[29] Thus, the amount of radio resources used for control signaling can be dynamically adjusted to match the instantaneous traffic situation. For a small number of users being scheduled in a subframe, the required amount of control signaling is small and a larger part of the subframe can be used for data transmission (larger data region).

The maximum size of the control region is normally three OFDM symbols (four in the case of narrow cell bandwidths), as mentioned previously. However, there are a few exceptions to this rule. When operating in TDD mode, the control region in subframes one and six is restricted to at most two OFDM symbols since, for TDD, the primary synchronization signal (see Chapter 14) occupies the third OFDM symbol in those subframes. Similarly, for MBSFN subframes (see Chapter 9), the control region is restricted to a maximum of two OFDM symbols.

The reason for transmitting the control signaling at the beginning of the subframe is to allow for terminals to decode downlink scheduling assignments as early as possible. Processing of the data region—that is, demodulation and decoding of the DL-SCH transmission—can then begin before the end of the subframe. This reduces the delay in the

[29]The start of the data region is semi-statically configured when using cross-carrier scheduling (see Section 10.4.1).

Copyright Elsevier 2024

DL-SCH decoding and thus the overall downlink transmission delay. Furthermore, by transmitting the L1/L2 control channel at the beginning of the subframe—that is, by allowing for early decoding of the L1/L2 control information—terminals that are not scheduled in the subframe may power down the receiver circuitry for a part of the subframe, allowing for a reduced terminal power consumption.

The downlink L1/L2 control signaling consists of five different physical-channel types, the first three located in the control region and the last two in the data region of a subframe:

- The *Physical Control Format Indicator Channel* (PCFICH), informing the terminal about the size of the control region (one, two, or three OFDM symbols). There is one and only one PCFICH on each component carrier or, equivalently, in each cell.
- The *Physical Hybrid-ARQ Indicator Channel* (PHICH), used to signal hybrid-ARQ acknowledgements in response to uplink UL-SCH transmissions. Multiple PHICHs can exist in each cell.
- The *Physical Downlink Control Channel* (PDCCH), used to signal downlink scheduling assignments, uplink scheduling grants, or power-control commands. Each PDCCH typically carries signalling for a single terminal, but can also be used to address a group of terminals. Multiple PDCCHs can exist in each cell.
- The *Enhanced Physical Downlink Control Channel* (EPDCCH), introduced in release 11 to support DM-RS-based control signalling reception and carrying similar types of information as the PDCCH. However, in contrast to the PDCCH, the EPDCCH is located in the data region. Also, the EPDCCH can be subject to non-codebook-based precoding.
- The *Relay Physical Downlink Control Channel* (R-PDCCH), introduced in release 10 to support relaying. A detailed discussion can be found in Chapter 18 in conjunction with the overall description of relays; at this stage it suffices to note that the R-PDCCH is transmitted in the data region.

In the following sections, the PCFICH, PHICH, PDCCH, and EPDCCH are described in detail.

10.4.1 Physical control format indicator channel

The PCFICH indicates the instantaneous size of the control region in terms of the number of OFDM symbols—that is, indirectly where in the subframe the data region starts. Correct decoding of the PCFICH information is thus essential. If the PCFICH is incorrectly decoded, the terminal will neither know how to process the control channels nor where the data region starts for the corresponding subframe.[30] The PCFICH consists of two bits of information, corresponding to the three[31]control-region sizes of one, two, or three OFDM symbols (two,

[30]Theoretically, the terminal could blindly try to decode all possible control channel formats and, from which format that was correctly decoded, deduce the starting position of the data region, but this can be a very complex procedure.

[31]The fourth combination is reserved for future use.

Copyright Elsevier 2024

three or four for narrow bandwidths), which are coded into a 32-bit codeword. The coded bits are scrambled with a cell- and subframe-specific scrambling code to randomize inter-cell interference, QPSK modulated, and mapped to 16 resource elements. As the size of the control region is unknown until the PCFICH is decoded, the PCFICH is always mapped to the first OFDM symbol of each subframe.

The mapping of the PCFICH to resource elements in the first OFDM symbol in the subframe is done in groups of four resource elements, with the four groups being well separated in frequency to obtain good diversity. Furthermore, to avoid collisions between PCFICH transmissions in neighboring cells, the location of the four groups in the frequency domain depends on the physical-layer cell identity.

The transmission power of the PCFICH is under control of the eNodeB. If necessary for coverage in a certain cell, the power of the PCFICH can be set higher than for other channels by "borrowing" power from, for example, simultaneously transmitted PDCCHs. Obviously, increasing the power of the PCFICH to improve the performance in an interference-limited system depends on the neighboring cells not increasing their transmit power on the interfering resource elements. Otherwise, the interference would increase as much as the signal power, implying no gain in received SIR. However, as the PCFICH-to-resource-element mapping depends on the cell identity, the probability of (partial) collisions with PCFICH in neighboring cells in synchronized networks is reduced, thereby improving the performance of PCFICH power boosting as a tool to control the error rate.

The overall PCFICH processing is illustrated in Figure 10.21.

To describe the mapping of the PCFICH, and L1/L2 control signaling in general, to resource elements, some terminology is required. As mentioned above, the mapping is specified in terms of groups of four resource elements, so-called *resource-element groups*. To each resource-element group, a *symbol quadruplet* consisting of four (QPSK) symbols is mapped. The main motivation behind this, instead of simply mapping the symbols one by one, is the support of transmit diversity. As discussed in Section 10.3, transmit diversity with

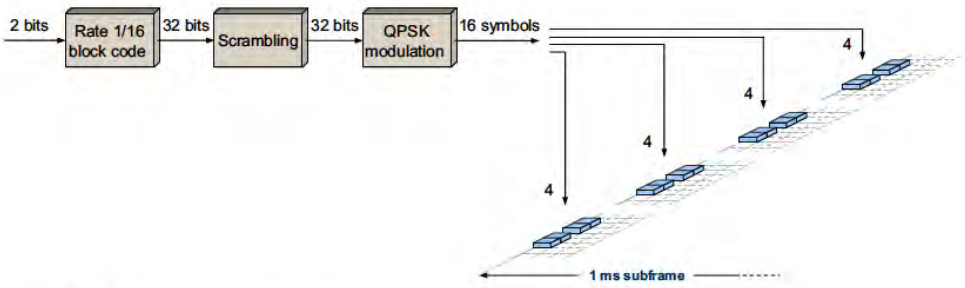


FIGURE 10.21

Overview of the PCFICH processing

Copyright Elsevier 2024

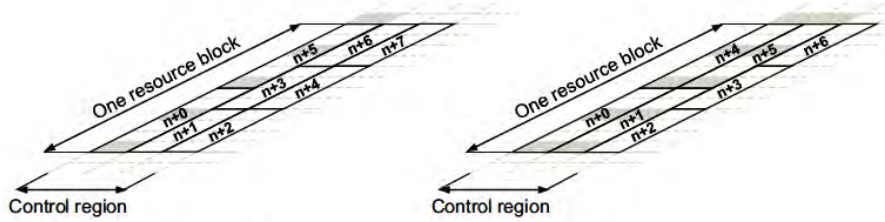


FIGURE 10.22

Numbering of resource-element groups in the control region (assuming a size of three OFDM symbols)

up to four antenna ports is specified for L1/L2 control signaling. Transmit diversity for four antenna ports is specified in terms of groups of four symbols (resource elements) and, consequently, the L1/L2 control-channel processing is also defined in terms of symbol quadruplets.

The definition of the resource-element groups assumes that reference symbols corresponding to two antenna ports are present in the first OFDM symbol, regardless of the actual number of antenna ports configured in the cell. This simplifies the definition and reduces the number of different structures to handle. Thus, as illustrated in Figure 10.22, in the first OFDM symbol there are two resource-element groups per resource block, as every third resource element is reserved for reference signals (or non-used resource elements corresponding to reference symbols on the other antenna port). As also illustrated in Figure 10.22, in the second OFDM symbol (if part of the control region) there are two or three resource-element groups depending on the number of antenna ports configured. Finally, in the third OFDM symbol (if part of the control region) there are always three resource-element groups per resource block. Figure 10.22 also illustrates how resource-element groups are numbered in a time-first manner within the size of the control region.

Returning to the PCFICH, four resource-element groups are used for the transmission of the 16 QPSK symbols. To obtain good frequency diversity the resource-element groups should be well spread in frequency and cover the full downlink cell bandwidth. Therefore, the four resource-element groups are separated by one-fourth of the downlink cell bandwidth in the frequency domain, with the starting position given by physical-layer cell identity. This is illustrated in Figure 10.23, where the PCFICH mapping to the first OFDM symbol in a subframe is shown for three different physical-layer cell identities in the case of a downlink cell bandwidth of eight resource blocks. As seen in the figure, the PCFICH mapping depends on the physical-layer cell identity to reduce the risk of inter-cell PCFICH collisions. The cell-specific shifts of the reference symbols, described in Section 10.2.1, are also seen in the figure.

For terminals capable of carrier aggregation, there is one PCFICH per component carrier. Independent signaling of the control-region size on the different component carriers is used,

Copyright Elsevier 2024

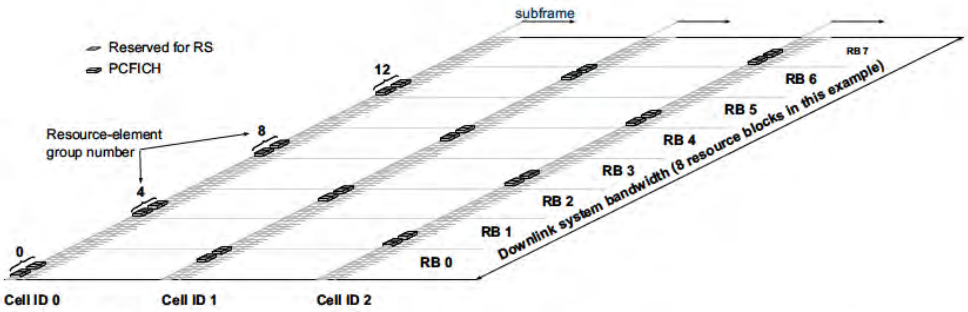


FIGURE 10.23

Example of PCFICH mapping in the first OFDM symbol for three different physical-layer cell identities

implying that the control region may be of different size on different component carriers. Hence, in principle the terminal needs to receive the PCFICH on each of the component carriers it is scheduled upon. Furthermore, as different component carriers may have different physical-layer cell identities, the location and scrambling may differ across component carriers.

If cross-carrier scheduling is used—that is, control signaling related to a certain PDSCH transmission is transmitted on a component carrier other than the PDSCH itself—the terminal needs to know the starting position for the data region on the carrier upon which the PDSCH is transmitted. Using the PCFICH on the component carrier carrying the PDSCH would be possible in principle, although it would increase the probability of incorrectly decoding the PDSCH since there are two PCFICH instances, one for the PDCCH decoding on one component carrier and one for PDSCH reception on the other component carrier. This would be problematic, especially since one use of cross-carrier scheduling is enhanced support of heterogeneous deployments (see Chapter 16), where some of the component carriers may be subject to strong interference. Therefore, for cross-carrier scheduled transmissions, the start of the data region is not obtained from the PCFICH on that component carrier, but is configured on a semi-static basis. The semi-statically configured value may differ from the value signaled on the PCFICH on the component carrier carrying the PDSCH transmission.

10.4.2 Physical hybrid-ARQ indicator channel

The PHICH is used for transmission of hybrid-ARQ acknowledgements in response to UL-SCH transmission. There is one PHICH transmitted per received transport block and TTI—that is, when uplink spatial multiplexing is used on a component carrier, two PHICHs are used to acknowledge the transmission, one per transport block.

Copyright Elsevier 2024

For proper operation of the hybrid-ARQ protocol and to avoid spurious uplink transmissions, the error rate of the PHICH should be sufficiently low. The operating point of the PHICH is not specified and is up to the network operator to decide upon, but typically ACK-to-NAK and NAK-to-ACK error rates of the order of 10^{-2} and 10^{-3}–10^{-4} respectively are targeted. The reason for the asymmetric error rates is that an NAK-to-ACK error would imply loss of a transport block at the MAC level, a loss that has to be recovered by RLC retransmissions with the associated delays, while an ACK-to-NAK error only implies an unnecessary retransmission of an already correctly decoded transport block. To meet these error-rate targets without excessive power, it is beneficial to control the PHICH transmission power as a function of the radio-channel quality of the terminal to which the PHICH is directed. This has influenced the design of the PHICH structure.

In principle, a PHICH could be mapped to a set of resource elements exclusively used by this PHICH. However, taking the dynamic PHICH power setting into account, this could result in significant variations in transmission power between resource elements, which can be challenging from an RF implementation perspective. Therefore, it is preferable to spread each PHICH on multiple resource elements to reduce the power differences while at the same time providing the energy necessary for accurate reception. To fulfill this, a structure where several PHICHs are code multiplexed on to a set of resource elements is used in LTE. The hybrid-ARQ acknowledgement (one single bit of information per transport block) is repeated three times, followed by BPSK modulation on either the I or the Q branch and spreading with a length-four orthogonal sequence. A set of PHICHs transmitted on the same set of resource elements is called a PHICH group, where a PHICH group consists of eight PHICHs in the case of normal cyclic prefix. An individual PHICH can thus be uniquely represented by a single number from which the number of the PHICH group, the number of the orthogonal sequence within the group, and the branch, I or Q, can be derived.

For extended cyclic prefix, which is typically used in highly time-dispersive environments, the radio channel may not be flat over the frequency spanned by a length-four sequence. A non-flat channel would negatively impact the orthogonality between the sequences. Hence, for extended cyclic prefix, orthogonal sequences of length two are used for spreading, implying only four PHICHs per PHICH group. However, the general structure remains the same as for the normal cyclic prefix.

After forming the composite signal representing the PHICHs in a group, cell-specific scrambling is applied and the 12 scrambled symbols are mapped to three resource-element groups. Similarly to the other L1/L2 control channels, the mapping is described using resource-element groups to be compatible with the transmit diversity schemes defined for LTE.

The overall PHICH processing is illustrated in Figure 10.24.

The requirements on the mapping of PHICH groups to resource elements are similar to those for the PCFICH, namely to obtain good frequency diversity and to avoid collisions between neighboring cells in synchronized networks. Hence, each PHICH group is mapped to

Copyright Elsevier 2024

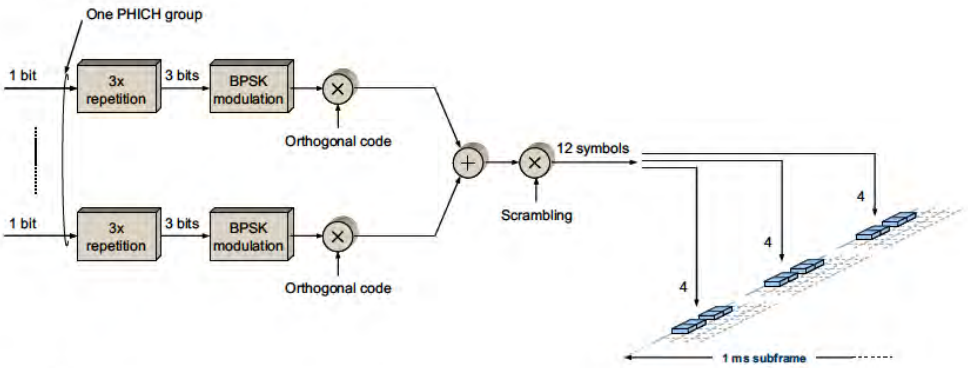


FIGURE 10.24

PHICH structure

three resource-element groups, separated by approximately one-third of the downlink cell bandwidth. In the first OFDM symbol in the control region, resources are first allocated to the PCFICH, the PHICHs are mapped to resource elements not used by the PCFICH, and finally, as will be discussed further below, the PDCCHs are mapped to the remaining resource elements.

Typically, the PHICH is transmitted in the first OFDM symbol only, which allows the terminal to attempt to decode the PHICH even if it failed decoding of the PCFICH. This is advantageous as the error requirements on the PHICH typically are stricter than for PCFICH. However, in some propagation environments, having a PHICH duration of a single OFDM symbol would unnecessarily restrict the coverage. To alleviate this, it is possible to semi-statically configure a PHICH duration of three OFDM symbols. In this case, the control region is three OFDM symbols long in all subframes to fulfill the general principle of separating the control region from data in the time domain only. The value transmitted on the PCFICH will be fixed (and can be ignored) in this case. For narrow bandwidths, where the control region can be up to four OFDM symbols long, there is still a need to use the PCFICH to differentiate between a size-three and size-four control region.

The PHICH configuration is part of the system information transmitted on the PBCH; one bit indicates whether the duration is one or three OFDM symbols, and two bits indicate the amount of resources in the control region reserved for PHICHs, expressed as a fraction of the downlink cell bandwidth in terms of resource blocks. Having the possibility to configure the amount of PHICH resources is useful as the PHICH capacity depends on, for example, whether the network uses multi-user MIMO or not. The PHICH configuration must reside on the PBCH, as it needs to be known in order to properly process the PDCCHs for reception of the part of the system information on the DL-SCH. For TDD, the PHICH information provided on the PBCH is not sufficient for the terminal to know the exact set of

Copyright Elsevier 2024

resources used by PHICH, as there is also a dependency on the uplink/downlink allocation, provided as part of the system information transmitted on the PDSCH. In order to receive the system information on the DL-SCH, which contains the uplink/downlink allocation, the terminal therefore has to blindly process the PDCCHs under different PHICH configuration hypotheses.

In order to minimize the overhead and not introduce any additional signaling in the uplink grants, the PHICH that the terminal will expect the hybrid-ARQ acknowledgement upon is derived from the number of the first resource block upon which the corresponding uplink PUSCH transmission occurred. This principle is also compatible with semi-persistently scheduled transmission (see Chapter 13) as well as retransmissions. In addition, the resources used for a particular PHICH further depend on the reference-signal phase rotation signaled as part of the uplink grant (see Section 10.4.7). In this way, multiple terminals scheduled on the same set of resources using multi-user MIMO will use different PHICH resources, as their reference signals are assigned different phase rotations through the corresponding field in the uplink grant. For spatial multiplexing, where two PHICH resources are needed, the second PHICH uses the same principle as the first, but to ensure that different PHICHs are used for the two transport blocks, the resource for the second PHICH is derived not from the first but from the second resource block upon which the PUSCH was transmitted.[32]

This association is also used in the case of carrier aggregation. As a general principle, LTE transmits the PHICH on the same component carrier that was used for the grantscheduling the corresponding uplink data transmission. Not only is this principle general in the sense that it can handle symmetric as well as asymmetric carrier aggregation, it is also beneficial from a terminal power consumption perspective as the terminal only need monitor the component carriers it monitors for uplink scheduling grants (especially as the PDCCH may override the PHICH to support adaptive retransmissions, as discussed in Chapter 12). For the case when no cross-carrier scheduling is used—that is, each uplink component carrier is scheduled on its corresponding downlink component carrier—different uplink component carriers will by definition have different PHICH resources. With cross-carrier scheduling, on the other hand, transmissions on multiple uplink component carriers may need to be acknowledged on a single downlink component carrier, as illustrated in Figure 10.25. Avoiding PHICH collisions in this case is up to the scheduler by ensuring that different reference-signal phase rotations or different resource-block starting positions are used for the different uplink component carriers. For semi-persistent scheduling the reference-signal phase rotation is always set to zero, but since semi-persistent scheduling is supported on the primary component carrier only, there is no risk of collisions between component carriers.

[32]In essence, this implies that uplink spatial multiplexing must use at least two resource blocks in the frequency domain.

Copyright Elsevier 2024

Samsung Exhibit 1017, Page 232 of 538
Samsung Electronics Co., Ltd. v. Asus Technology Licensing Inc.
IPR2024-01217

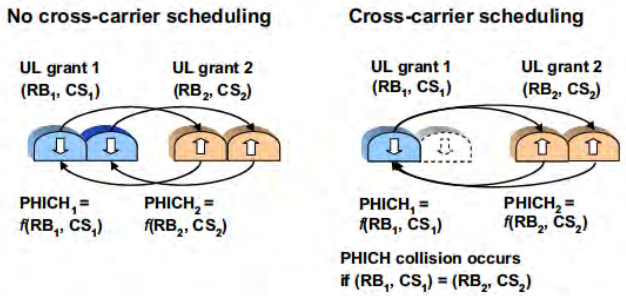


FIGURE 10.25

PHICH association

10.4.3 Physical downlink control channel

The PDCCH is used to carry downlink control information (DCI) such as scheduling decisions and power-control commands. More specifically, the DCI includes:

- Downlink scheduling assignments, including PDSCH resource indication, transport format, hybrid-ARQ information, and control information related to spatial multiplexing (if applicable). A downlink scheduling assignment also includes a command for power control of the PUCCH used for transmission of hybrid-ARQ acknowledgements in response to downlink scheduling assignments.
- Uplink scheduling grants, including PUSCH resource indication, transport format, and hybrid-ARQ-related information. An uplink scheduling grant also includes a command for power control of the PUSCH.
- Power-control commands for a set of terminals as a complement to the commands included in the scheduling assignments/grants.

The different types of control information, both between the groups above as well as within the groups, correspond to different DCI message sizes. For example, supporting spatial multiplexing with non-contiguous allocation of resource blocks in the frequency domain requires a larger scheduling message in comparison with an uplink grant allowing for frequency-contiguous allocations only. The DCI is therefore categorized into different *DCI formats*, where a format corresponds to a certain message size and usage. The DCI formats are summarized in Table 10.4, including the size for an example of 20 MHz FDD operation with 2 Tx antennas at the base station and no carrier aggregation. The actual message size depends on, among other factors, the cell bandwidth as, for larger bandwidths, a larger number of bits is required to indicate the resource-block allocation. The number of cell-specific reference signals in the cell and whether cross-carrier scheduling is configured or not will also affect the absolute size of most DCI formats. Hence, a given DCI format may

Copyright Elsevier 2024

Table 10.4 DCI Formats

	DCI Format	Size (Bits)	Usage
Uplink	0	45	Uplink scheduling grant
	4	53	Uplink scheduling grant with spatial multiplexing
Downlink	1C	31	Special purpose compact assignment
	1A	45	Contiguous allocations only
	1B	46	Codebook-based beam-forming using CRS
	1D	46	Multi-user MIMO using CRS
	1	55	Flexible allocations
	2A	64	Open-loop spatial multiplexing using CRS
	2B	64	Dual-layer transmission using DM-RS (TM8)
	2C	66	Multi-layer transmission using DM-RS (TM9)
	2D	68	Multi-layer transmission using DM-RS (TM10)
	2	67	Closed-loop spatial multiplexing using CRS
Special	3, 3A	45	Power control commands

have different sizes depending on the overall configuration of the cell. This will be discussed further below; at this stage it suffices to note that formats 0, 1A, 3, and 3A have the same message size.[33]

One PDCCH carries one DCI message with one of the formats above. To support different radio-channel conditions, link adaptation can be used, where the code rate (and transmission power) of the PDCCH is selected to match the radio-channel conditions. As multiple terminals can be scheduled simultaneously, on both downlink and uplink, there must be a possibility to transmit multiple scheduling messages within each subframe. Each scheduling message is transmitted on a separate PDCCH, and consequently there are typically multiple simultaneous PDCCH transmissions within each cell. A terminal may also receive ***multiple*** DCI messages in the same subframe (on different PDCCHs), for example, if it is scheduled simultaneously in uplink and downlink.

[33]The smaller of DCI formats 0 and 1A is padded to ensure the same payload size. Which of 0 and 1A is padded depends on the uplink and downlink cell bandwidths; in the case of identical uplink and downlink bandwidths in a cell there is a single bit of padding in format 0.

Copyright Elsevier 2024

For carrier aggregation, scheduling assignments/grants are transmitted individually per component carrier. PDCCH transmission for carrier aggregation is further elaborated upon in Section 10.4.8.

Having introduced the concept of DCI formats, the transmission of the DCI message on a PDCCH can be described. The processing of downlink control signaling is illustrated in Figure 10.26. A CRC is attached to each DCI message payload. The identity of the terminal (or terminals) addressed—that is, the RNTI—is included in the CRC calculation and not explicitly transmitted. Depending on the purpose of the DCI message (unicast data transmission, power-control command, random-access response, etc.), different RNTIs are used; for normal unicast data transmission, the terminal-specific C-RNTI is used.

Upon reception of DCI, the terminal will check the CRC using its set of assigned RNTIs. If the CRC checks, the message is declared to be correctly received and intended for the terminal. Thus,

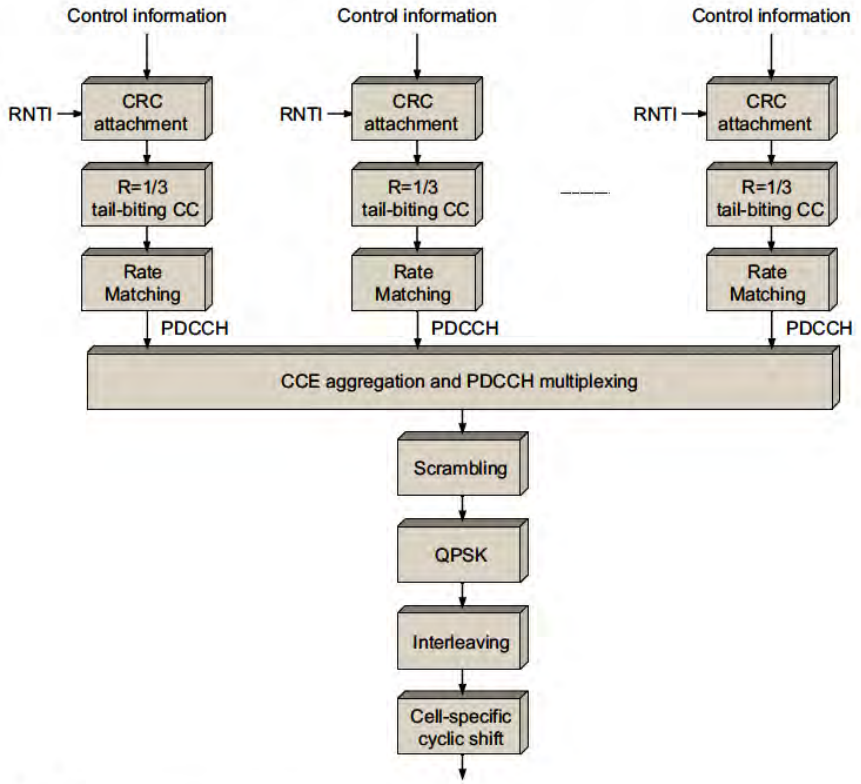


FIGURE 10.26

Processing of L1/L2 control signaling

Copyright Elsevier 2024

the identity of the terminal that is supposed to receive the DCI message is implicitly encoded in the CRC and not explicitly transmitted. This reduces the amount of bits necessary to transmit on the PDCCH as, from a terminal point of view, there is no difference between a corrupt message whose CRC will not check and a message intended for another terminal.

After CRC attachment, the bits are coded with a rate-1/3 tail-biting convolutional code and rate-matched to fit the amount of resources used for PDCCH transmission. Tail-biting convolutional coding is similar to conventional convolutional coding with the exception that no tail bits are used. Instead, the convolutional encoder is initialized with the last bits of the message prior to the encoding process. Thus, the starting and ending states in the trellis in an MLSE (Viterbi) decoder are identical.

After the PDCCHs to be transmitted in a given subframe have been allocated to the desired resource elements (the details of which are given below), the sequence of bits corresponding to all the PDCCH resource elements to be transmitted in the subframe, including the unused resource elements, is scrambled by a cell- and subframe-specific scrambling sequence to randomize inter-cell interference, followed by QPSK modulation and mapping to resource elements.

To allow for simple yet efficient processing of the control channels in the terminal, the mapping of PDCCHs to resource elements is subject to a certain structure. This structure is based on so-called *Control-Channel Elements* (CCEs), which in essence is a convenient name for a set of 36 useful resource elements (nine resource-element groups as defined in Section 10.4.1). The number of CCEs, one, two, four, or eight, required for a certain PDCCH depends on the payload size of the control information (DCI payload) and the channel-coding rate. This is used to realize link adaptation for the PDCCH; if the channel conditions for the terminal to which the PDCCH is intended are disadvantageous, a larger number of CCEs needs to be used compared to the case of advantageous channel conditions. The number of CCEs used for a PDCCH is also referred to as the *aggregation level*.

The number of CCEs available for PDCCHs depends on the size of the control region, the cell bandwidth, the number of downlink antenna ports, and the amount of resources occupied by PHICH. The size of the control region can vary dynamically from subframe to subframe as indicated by the PCFICH, whereas the other quantities are semi-statically configured. The CCEs available for PDCCH transmission can be numbered from zero and upward, as illustrated in Figure 10.27. A specific PDCCH can thus be identified by the numbers of the corresponding CCEs in the control region.

As the number of CCEs for each of the PDCCHs may vary and is not signaled, the terminal has to blindly determine the number of CCEs used for the PDCCH it is addressed upon. To reduce the complexity of this process somewhat, certain restrictions on the aggregation of contiguous CCEs have been specified. For example, an aggregation of eight CCEs can only start on CCE numbers evenly divisible by 8, as illustrated in Figure 10.27. The same principle is applied to the other aggregation levels. Furthermore, some combinations of DCI formats and CCE aggregations that result in excessively high channel-coding rates are not supported.

Copyright Elsevier 2024

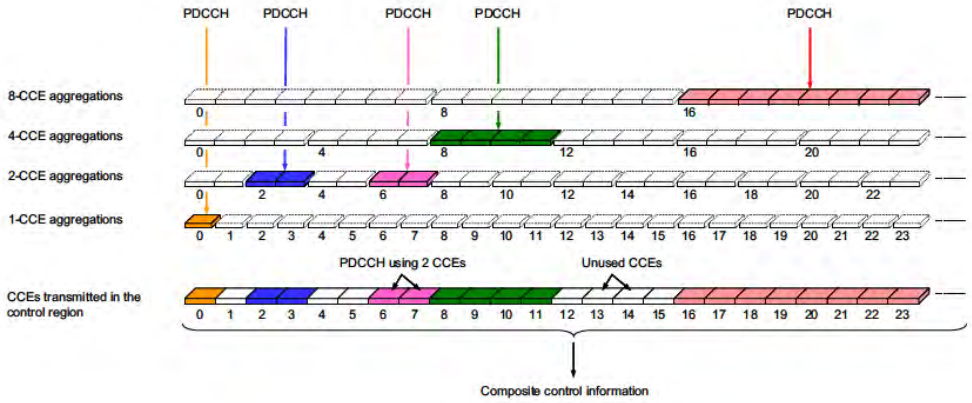


FIGURE 10.27

CCE aggregation and PDCCH multiplexing

The sequence of CCEs should match the amount of resources available for PDCCH transmission in a given subframe—that is, the number of CCEs varies according to the value transmitted on the PCFICH. In many cases, not all the PDCCHs that can be transmitted in the control region are used. Nevertheless, unused PDCCHs are part of the interleaving and mapping process in the same way as any other PDCCH. At the terminal, the CRC will not check for those "dummy" PDCCHs. Preferably, the transmission power is set to zero for those unused PDCCHs; the power can be used by other control channels.

The mapping of the modulated composite control information is, for the same reason as for PCFICH and PHICH, described in terms of symbol quadruplets being mapped to resource-element groups. Thus, the first step of the mapping stage is to group the QPSK symbols into symbol quadruplets, each consisting of four consecutive QPSK symbols. In principle, the sequence of quadruplets could be mapped directly to the resource elements in sequential order. However, this would not exploit all the frequency diversity available in the channel, and diversity is important for good performance. Furthermore, if the same CCE-to-resource-element mapping is used in all neighboring cells, a given PDCCH will persistently collide with one and the same PDCCH in the neighboring cells assuming a fixed PDCCH format and inter-cell synchronization. In practice, the number of CCEs per PDCCH varies in the cell as a function of the scheduling decisions, which gives some randomization to the interference, but further randomization is desirable to obtain a robust control-channel design. Therefore, the sequence of quadruplets is first interleaved using a block interleaver to allow exploitation of the frequency diversity, followed by a cell-specific cyclic shift to randomize the interference between neighboring cells. The output from the cell-specific shift is mapped to resource-element groups in a time-first manner, as illustrated in Figure 10.22,

Copyright Elsevier 2024

skipping resource-element groups used for PCFICH and PHICH. Time-first mapping preserves the interleaving properties; with frequency-first over multiple OFDM symbols, resource-element groups that are spread far apart after the interleaving process may end up close in frequency, although on different OFDM symbols.

The interleaving operation described above, in addition to enabling exploitation of the frequency diversity and randomizing the inter-cell interference, also serves the purpose of ensuring that each CCE spans virtually all the OFDM symbols in the control region. This is beneficial for coverage as it allows flexible power balancing between the PDCCHs to ensure good performance for each of the terminals addressed. In principle, the energy available in the OFDM symbols in the control region can be balanced arbitrarily between the PDCCHs. The alternative of restricting each PDCCH to a single OFDM symbol would imply that power cannot be shared between PDCCHs in different OFDM symbols.

Similarly to the PCFICH, the transmission power of each PDCCH is under the control of the eNodeB. Power adjustments can therefore be used as a complementary link adaptation mechanism in addition to adjusting the code rate. Relying on power adjustments alone might seem a tempting solution but, although possible in principle, it can result in relatively large power differences between resource elements. This may have implications on the RF implementation and may violate the out-of-band emission masks specified. Hence, to keep the power differences between the resource elements reasonable, link adaptation through adjusting the channel code rate, or equivalently the number of CCEs aggregated for a PDCCH, is necessary. Furthermore, lowering the code rate is generally more efficient than increasing the transmission power. The two mechanisms for link adaptation, power adjustments and different code rates, complement each other.

To summarize and to illustrate the mapping of PDCCHs to resource elements in the control region, consider the example shown in Figure 10.28. In this example, the size of the control region in the subframe considered equals three OFDM symbols. Two downlink antenna ports are configured (but, as explained above, the mapping would be identical in the case of a single antenna port). One PHICH group is configured and three resource-element groups are therefore used by the PHICHs. The cell identity is assumed to be identical to zero in this case.

The mapping can then be understood as follows: First, the PCFICH is mapped to four resource-element groups, followed by allocating the resource-element groups required for the PHICHs. The resource-element groups left after the PCFICH and PHICHs are used for the different PDCCHs in the system. In this particular example, one PDCCH is using CCE numbers 0 and 1, while another PDCCH is using CCE number 4. Consequently, there is a relatively large amount of unused resource-element groups in this example; either they can be used for additional PDCCHs or the power otherwise used for the unused CCEs could be allocated to the PDCCHs in use (as long as the power difference between resource elements is kept within the limits set by the RF requirements). Furthermore, depending on the inter-cell interference situation, fractional loading of the control region may be desirable, implying that some CCEs are left unused to reduce the average inter-cell interference.

Copyright Elsevier 2024

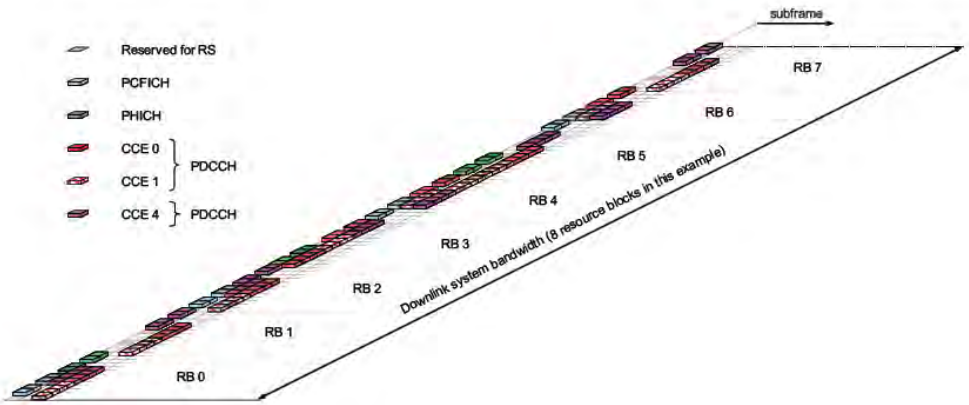


FIGURE 10.28

Example of mapping of PCFICH, PHICH, and PDCCH

10.4.4 Enhanced physical downlink control channel

In release 11, a complementary control channel was introduced—the *Enhanced Physical Downlink Control Channel* (EPDCCH)—primarily differing from the PDCCH in terms of precoding and mapping to physical resources. The reasons for introducing the EPDCCH are twofold:

- To enable frequency-domain scheduling and interference coordination also for control signaling.
- To enable DM-RS-based reception for the control signaling.

Unlike the PDCCH, which is transmitted in the control region and spans the full system bandwidth, the EPDCCH is transmitted in the data region and normally spans only a small portion of the overall bandwidth (see Figure 10.29). Hence, as it is possible to control in which part of the overall spectrum an EPDCCH is transmitted, it is possible not only to benefit

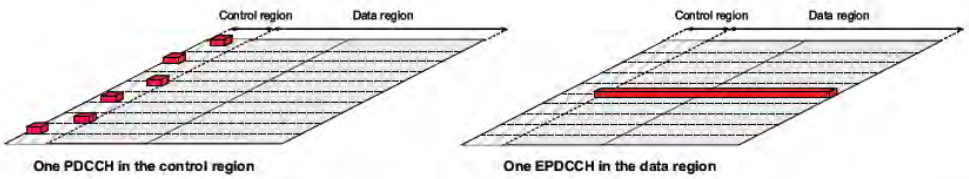


FIGURE 10.29

Principal illustration of PDCCH mapping (left) and EPDCCH mapping (right)

Copyright Elsevier 2024

from frequency-selective scheduling for control channels, but also to implement various forms of inter-cell interference coordination schemes in the frequency domain. This can be very useful in, for example, heterogeneous deployments (see Chapter 16). Furthermore, as the EPDCCH uses DM-RS for demodulation in contrast to the CRS-based reception of the PDCCH, any precoding operations are therefore transparent to the terminal. The EPDCCH can therefore be seen as a prerequisite for CoMP (see Chapter 15) as well as more advanced antenna solutions in general for which it is beneficial for the network to have the freedom of changing the precoding without explicitly informing the terminal.

In contrast to the PDCCH, the EPDCCH decoding result is not available until the end of the subframe, which leaves less processing time for the PDSCH. Hence, for the largest DL-SCH payloads, the terminal may need a slightly faster PDSCH decoder compared to the PDCCH case to meet the hybrid-ARQ timing.

With a few exceptions, the EPDCCH generally supports the same DCI formats as the PDCCH (see Table 10.4) and can thus be used for scheduling all the different transmission modes in both uplink and downlink. However, DCI formats 3/3A, used for uplink power control of multiple terminals, and format 1C, used for scheduling system information to multiple terminals, are not supported on the EPDCCH for reasons discussed in conjunction with search spaces in Section 10.4.5.

The EPDCCH processing, illustrated in Figure 10.30, is virtually identical to that of a PDCCH apart from precoding and the resource-element mapping. The identity of the terminal addressed—that is, the RNTI—is implicitly encoded in the CRC and not explicitly transmitted. After CRC attachment, the bits are coded with a rate-1/3 tail-biting convolutional code and rate-matched to fit the amount of resources available for the EPDCCH in question. The coded and rate-matched bits are scrambled with a terminal and EPDCCH-set specific scrambling sequence (EPDCCH sets are described below) and QPSK-modulated before being mapped to the physical resources. Link adaptation is supported by adjusting the number of control-channel resources utilized by an encoded EPDCCH message, similar to the PDCCH.

To describe the mapping of EPDCCH to physical resources, it is worthwhile to view this from a terminal perspective. Each terminal using EPDCCH is configured with one or two sets of physical resource blocks where EPDCCH transmission to that terminal may occur. Each set consists of two, four, or eight physical resource-block pairs, and the two sets may be of different size. There is full flexibility in the location of the resource-block pairs across the full downlink system bandwidth (the location is configured by higher layers using a combinatorial index) and the resource-block pairs do not have to be contiguous in frequency. It is important to understand that these sets are defined from a *terminal* perspective and only indicates where a terminal *may* receive EPDCCH transmissions. A resource-block pair not used for EPDCCH transmission to a particular terminal in a certain subframe can still be used for data transmission, either to the same terminal or to another terminal, despite being part of one of the two EPDCCH sets.

An EPDCCH set is configured as being either *localized* or *distributed*. For a localized set, a single EPDCCH is mapped to one (or a few in case of the highest aggregation levels) physical

Copyright Elsevier 2024

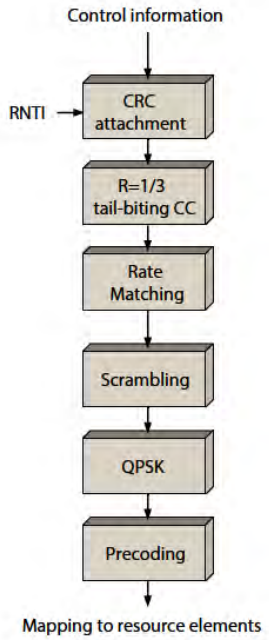


FIGURE 10.30

EPDCCH coding

resource-block pair to allow exploitation of frequency-domain scheduling gains. For a distributed set, on the other hand, a single EPDCCH is distributed over multiple physical resource-block pairs with the intention of providing robustness for situations in which there is limited or no channel-state information available at the transmitter.

Providing two EPDCCH sets instead of just one is beneficial for several reasons. For example, one set can be configured for localized transmission in order to benefit from frequency-dependent scheduling, and the other set can be configured for distributed transmission and act as a fallback in case the channel-state feedback becomes unreliable due to the terminal moving rapidly. It is also useful for supporting CoMP as discussed further below.

Inspired by the PDCCH, the EPDCCH mapping to physical resources is subject to a structure based on *enhanced control-channel elements* (ECCEs) and *enhanced resource-element groups* (EREGs). An EPDCCH is mapped to a number of ECCEs, where the number of ECCEs used for an EPDCCH is known as the *aggregation level* and is used to realize link adaptation. If the radio conditions are worse, or if the EPDCCH payload is large, a larger number of ECCEs is used than for small payloads and/or good radio conditions. An EPDCCH uses a consecutive set of ECCEs in which the number of ECCEs used for an EPDCCH can be

Copyright Elsevier 2024

1, 2, 4, 8, 16, or 32, although not all these values are possible for all configurations. Note though that the ECCE numbering is in the *logical* domain; consecutive ECCEs does not necessarily imply transmission on consecutive physical resource-block pairs.

Each ECCE in turn consists of four[34] EREGs, where an EREG in essence corresponds to nine[35] resource elements in one physical resource-block pair. To define an EREG, number all resource elements in a physical resource-block pair cyclically in a frequency-first manner from 0 to 15, excluding resource-elements used for DM-RS. EREG number i consists of all resource elements with number i in that physical resource-block pair, and there are thus 16 EREGs in one physical resource-block pair. However, note that not all nine resource elements in an EREG may be available for EPDCCH usage, as some resource elements are occupied by, for example, the PDCCH control region, cell-specific reference signals, or CSI reference signals. In Figure 10.31, an example with EREG 0 in one physical resource-block pair is shown.

The mapping of ECCEs to EREGs is different for localized and distributed transmission (see Figure 10.32).

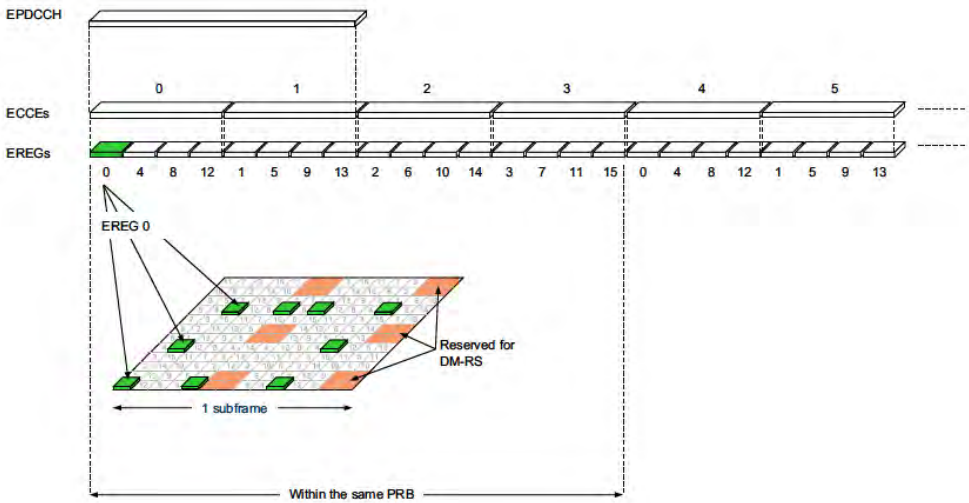


FIGURE 10.31

Example of relation between EPDCCH, ECCEs, and EREGs for localized mapping with aggregation level 2

[34]Four holds for normal cyclic prefix; for extended cyclic prefix, and some special subframe configurations in normal cyclic prefix, there are eight EREGs per ECCE.
[35]Eight in case of extended cyclic prefix.

Copyright Elsevier 2024

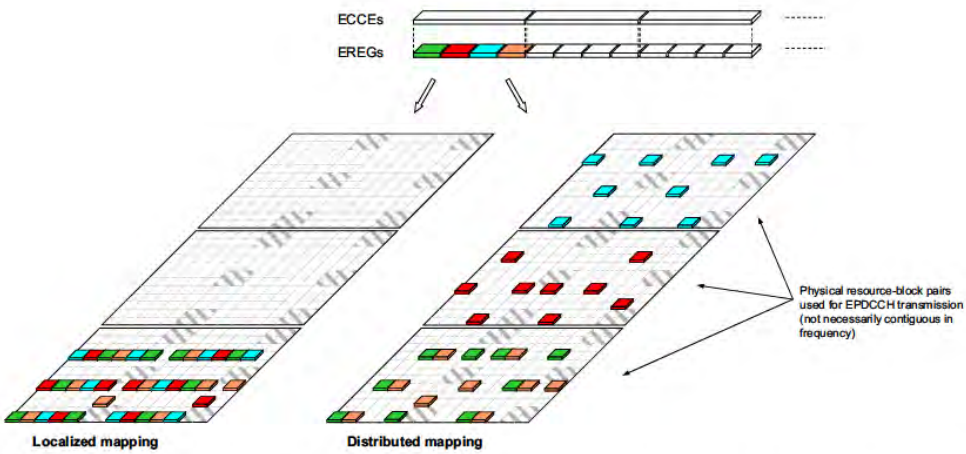


FIGURE 10.32

Example of localized and distributed mapping of ECCEs (three physical resource-block pairs for EPDCCH assumed in this example)

For localized transmission, the intention is to provide possibilities to select the physical resources and antenna precoding based on the instantaneous channel conditions. This is useful when exploiting channel-dependent scheduling or in conjunction with multi-antenna schemes such as CoMP (see Chapter 15). Hence, an ECCE is mapped to EREGs in the *same* physical resource-block pair, which allows the eNodeB to select the appropriate frequency region to which to transmit the EPDCCH. Only if one physical resource-block pair is not sufficient to carry one EPDDCH, that is, for the highest aggregation levels, a second physical resource-block pair is used. A single antenna port is used for transmission of the EPDCCH. The DM-RS associated with the EPDCCH transmission[36] is a function of the ECCE index and the C-RNTI. This is useful to support multi-user MIMO (see Section 10.3.5) in which multiple EPDCCHs intended for different spatially separated terminals are transmitted using the *same* time-frequency resources, but with *different* DM-RS sequences. Up to four different orthogonal DM-RS sequences are available, implying that up to four different terminals can be multiplexed in this manner.

For distributed transmission, the intention is to maximize diversity to provide diversity gain in situations in which it is not possible or desirable to exploit the instantaneous channel conditions. Consequently, an ECCE is mapped to EREGs in *different* physical resource-block pairs, where the physical resource-block pairs preferably are configured to be far apart in the frequency domain. To further increase the diversity order and exploit multiple transmission

[36]In the specifications, this is described as selecting *one* out of *four* antenna ports for transmission of the EPDCCH, where each antenna port has an associated orthogonal reference signal.

Copyright Elsevier 2024

antennas, the resource elements in each EREG are alternating between two antenna ports to provide spatial diversity.

Quasi co-location configuration is supported for the EPDCCH. In transmission mode 10, each EPDCCH set is associated with one of the four *PDSCH to RE mapping and quasi-co-location states* (see Section 9.3.1). This can be used to enable CoMP and dynamic point selection (see Chapter 15). Each EPDCCH set is associated with a certain transmission point, and dynamic point selection is achieved by selecting from which set to transmit a DCI message. Handling different sizes of the PDCCH control region and locations of cell-specific reference signals from the different transmission points is possible since these pieces of information are part of the four states.

Triggering of retransmissions is done in the same manner as for the PDCCH case, that is, by using the PHICH, which in essence is a very compact uplink grant for retransmissions. In principle an EPHICH could be thought of, but since each EPHICH would need its own set of DM-RS, the resulting structure would not be as compact as the PHICH in which multiple terminals share the same set of cell-specific reference signals. Hence, either the PHICH is used or, if the dependency on cell-specific reference signals is not desirable, an EPDCCH can be used to schedule a retransmission.

10.4.5 Blind decoding of PDCCHs and EPDCCHs

As described above, each PDCCH or EPDCCH supports multiple DCI formats and the format used is a priori unknown to the terminal. Therefore, the terminal needs to blindly detect the format of the (E)PDCCHs. The CCE and ECCE structures described in the previous section help in reducing the number of blind decoding attempts, but are not sufficient. Hence, it is required to have mechanisms to limit the number of CCE/ECCE aggregations that the terminal is supposed to monitor. Clearly, from a scheduling point of view, restrictions in the allowed aggregations are undesirable as they may influence the scheduling flexibility and require additional processing at the transmitter side. At the same time, requiring the terminal to monitor all possible CCE/ECCE aggregations, also for the larger cell bandwidths and EPDCCH set sizes, is not attractive from a terminal-complexity point of view. To impose as few restrictions as possible on the scheduler while at the same time limit the maximum number of blind decoding attempts in the terminal, LTE defines so-called *search spaces.* A search space is a set of candidate control channels formed by CCEs (or ECCEs) at a given aggregation level, which the terminal is supposed to attempt to decode. As there are multiple aggregation levels, a terminal has multiple search spaces. The search space concept is applied to both PDCCH and EPDCCH decoding, although there are differences between the two. In the following section, search spaces are described, starting with the case of PDCCHs on a single component carrier and later extended to EPDCCHs as well as multiple component carriers.

The PDCCH supports four different aggregation levels corresponding to one, two, four, and eight CCEs respectively. In each subframe, the terminals will attempt to decode all the

Copyright Elsevier 2024

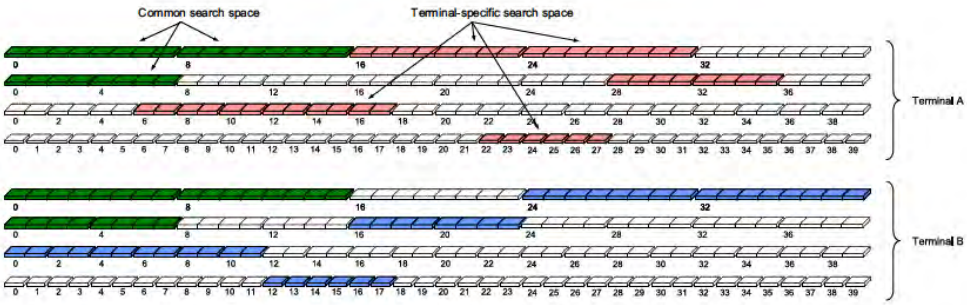


FIGURE 10.33

Principal illustration of PDCCH search spaces in two terminals

PDCCHs that can be formed from the CCEs in each of its search spaces. If the CRC checks, the content of the control channel is declared as valid for this terminal and the terminal processes the information (scheduling assignment, scheduling grants, etc.). The network can only address a terminal if the control information is transmitted on a PDCCH formed by the CCEs in one of the terminal's search spaces. For example, terminal A in Figure 10.33 cannot be addressed on a PDCCH starting at CCE number 20, whereas terminal B can. Furthermore, if terminal A is using CCEs 16–23, terminal B cannot be addressed on aggregation level 4 as all CCEs in its level-4 search space are blocked by the use for the other terminals. From this it can be intuitively understood that for efficient utilization of the CCEs in the system, the search spaces should differ between terminals. Each terminal in the system therefore has a *terminal-specific* search space at each aggregation level.

As the terminal-specific search space is typically smaller than the number of PDCCHs the network could transmit at the corresponding aggregation level, there must be a mechanism determining the set of CCEs in the terminal-specific search space for each aggregation level. One possibility would be to let the network configure the terminal-specific search space in each terminal. However, this would require explicit signaling to each of the terminals and possibly reconfiguration at handover. Instead, the terminal-specific search spaces for PDCCH are defined without explicit signaling through a function of the terminal identity and the subframe number. Dependence on the subframe number results in the terminal-specific search spaces being time varying, which helps resolve blocking between terminals. If a given terminal cannot be scheduled in a subframe as all the CCEs that the terminal is monitoring have already been used for scheduling other terminals in the same subframe, the time-varying definition of the terminal-specific search spaces is likely to resolve the blocking in the next subframe.

In several situations, there is a need to address a group of, or all, terminals in the system. One example is dynamic scheduling of system information; another is transmission of paging

Copyright Elsevier 2024

messages, both described in Chapter 14. Transmission of explicit power-control commands to a group of terminals is a third example. To allow multiple terminals to be addressed at the same time, LTE has defined *common search spaces* for the PDCCH in addition to the terminal-specific search spaces. A common search space is, as the name implies, common, and all terminals in the cell monitor the CCEs in the common search spaces for PDCCH control information. Although the motivation for the common search space is primarily transmission of various system messages, it can be used to schedule individual terminals as well. Thus, it can be used to resolve situations where scheduling of one terminal is blocked due to lack of available resources in the terminal-specific search space. Unlike unicast transmissions, where the transmission parameters of the control signaling can be tuned to match the channel conditions of a specific terminal, system messages typically need to reach the cell border. Consequently, the common search spaces are only defined for aggregation levels of four and eight CCEs and only for the smallest DCI formats, 0/1A/3/3A and 1C. There is no support for DCI formats with spatial multiplexing in the common search space. This helps to reduce the number of blind decoding attempts in the terminal used for monitoring the common search space.

Figure 10.33 illustrates the terminal-specific and common search spaces for PDCCH in two terminals in a certain subframe. The terminal-specific search spaces are different in the two terminals and will, as described above, vary from subframe to subframe. Furthermore, the terminal-specific search spaces partially overlap between the two terminals in this subframe (CCEs 24–31 on aggregation level 8) but, as the terminal-specific search space varies between subframes, the overlap in the next subframe is most likely different. There are 16 PDCCH candidates in the terminal-specific search spaces, mainly allocated to the smaller aggregation levels, and six PDCCH candidates in the common search spaces.

EPDCCH blind decoding in general follows the same principles as the PDCCH—that is, the terminal will attempt to decode all the EPDCCHs that can be formed from the ECCEs in each of its search spaces. However, only the terminal-specific search spaces are supported for EPDCCH. Thus, if a terminal has been configured to use the EPDCCH, it monitors the EPDCCH terminal-specific search spaces *instead of* the PDCCH terminal-specific search spaces. The common search spaces for PDCCH are monitored irrespective of whether an EPDCCH has been configured or not. The reason for not defining common search spaces for EPDCCH is that system information needs to be provided to all terminals, including those not supporting EPDCCH, and hence the PDCCH needs to be used. The lack of a common search space for the EPDCCH is also the reason why DCI formats 3/3A and 1C are not supported on the EPDCCH.

The terminal-specific EPDCCH search spaces for an EPDCCH set are randomly varying within the set of physical resource-block pairs configured for EPDCCH monitoring in a terminal using similar principles as the PDCCH. The two EPDCCH sets have independent pseudo-random sequences, initialized through RRC signaling. Having multiple EPDCCH sets is beneficial as it reduces the blocking probability. Compare with the PDCCH, where one PDCCH on aggregation level eight may block all the PDCCH candidates on aggregation level one.

Copyright Elsevier 2024

For a localized EPDCCH set, the EPDCCH candidates are spread over as many physical resource-block pairs as possible. The intention behind this is to allow the terminal to be addressed over a wide frequency range in order not to limit the benefits of channel-dependent scheduling of EPDCCH.

The number of blind decoding attempts is the same irrespective of whether the terminal is monitoring the EPDCCH or PDCCH search spaces. Thus, there are 16 EPDCCH candidates[37] to attempt to decode, and for each candidate two different DCI formats need to be considered. The 16 EPDCCH candidates are distributed across the two EPDDCH sets, roughly following the principles that the number of candidates is proportional to the number of physical resource-block pairs. Furthermore, for a localized set, more candidates are allocated to lower aggregation levels suitable for good channel conditions, motivated by the assumption that channel-dependent scheduling is used for localized transmission. For distributed transmission, the opposite holds, that is, more candidates are allocated to the higher aggregation levels than the lower ones to provide robustness as channel-dependent scheduling is typically not used for distributed transmission.

As mentioned above, the terminal is monitoring *either* the terminal-specific PDCCH search spaces *or* the terminal-specific EPDCCH search spaces. The basic principle is that the terminal is monitoring EPDCCHs in all subframes whenever EPDCCH support is enabled, except for the special subframe in configurations where there are no DM-RS supported and hence no possibility to receive the EPDCCH. However, to provide additional control of when the EPDCCH is monitored, it is possible to provide a bitmap to the terminal, indicating in which subframes it should monitor EPDCCHs and in which subframes it should monitor PDCCHs. One possible use case for this is PDSCH transmission in MBSFN subframes. In MBSFN subframes where the PMCH is transmitted, it is not possible to transmit the EPDCCH and the terminal needs to be addressed using the PDCCH, while in MBSFN subframes not used for PMCH transmission, the EPDCCH can be used.

Previously, the search spaces for PDCCH and EPDCCH in terms of control channel candidates were described. However, to determine whether a control channel candidate contains relevant downlink control information or not, the contents have to be decoded. If the CRC, which includes the terminal identity, checks, then the content of the control channel is declared as valid for this terminal, and the terminal processes the information (scheduling assignment, scheduling grants, etc.). Thus, for each control-channel candidate, the terminal needs to attempt to decode the contents once for each supported DCI format. The downlink DCI formats to decode in the terminal-specific search spaces depend on the *transmission mode* configured for the terminal. Transmission modes are described in Section 10.3.1 and, in principle, correspond to different multi-antenna configurations. As an example, there is no need to attempt to decode DCI format 2 when the terminal has not been configured for spatial multiplexing, which helps to reduce the number of blind decoding attempts. The DCI formats

[37]For a small number of physical resource-block pairs configured for EPDCCH there may be less than 16 EPDCCH candidates.

Copyright Elsevier 2024

Samsung Exhibit 1017, Page 247 of 538
Samsung Electronics Co., Ltd. v. Asus Technology Licensing Inc.
IPR2024-01217

Table 10.5 Downlink DCI Formats Monitored in Different Search Spaces for C-RNTI

Mode	Search Space: Common (PDCCH)	Search Space: Terminal-Specific (PDCCH or EPDCCH)		Description	Release
1	1A	1A	1	Single antenna transmission	Rel-8
2	1A	1A	1	Transmit diversity	
3	1A	1A	2A	Open-loop spatial multiplexing	
4	1A	1A	2	Closed-loop spatial multiplexing	
5	1A	1A	1D	Multi-user MIMO	
6	1A	1A	1B	Single layer codebook-based precoding	
7	1A	1A	1	Single-layer transmission using DM-RS	
8	1A	1A	2B	Dual-layer transmission using DM-RS	Rel-9
9	1A	1A	2C	Multi-layer transmission using DM-RS	Rel-10
10	1A	1A	2D	Multi-layer transmission using DM-RS	Rel-11

a terminal addressed using the C-RNTI should monitor as a function of the transmission mode, as listed in Table 10.5. Note that DCI format 1C is monitored in the common search spaces as well, although not with the C-RNTI identity. As seen from the table, there are two DCI formats to monitor in the terminal-specific search spaces and one in the common search spaces. In addition, a terminal also needs to monitor DCI format 1C in the common search space. Hence, with 16 PDCCH/EPDCCH candidates in the terminal-specific search spaces and six in the common search spaces, a terminal needs to execute $2 \cdot 16 + 2 \cdot 6 = 44$ blind-decoding attempts in each subframe. With uplink spatial multiplexing, introduced in release 10, an additional uplink DCI format needs to be monitored in the terminal-specific search spaces, increasing the number of blind decoding attempts to $3 \cdot 16 + 2 \cdot 6 = 60$. These numbers are for a single component carrier; in case of carrier aggregation, the number of blind decodes is increased further, as discussed in Section 10.4.8. Finally, note that a certain terminal in some circumstances may be addressed with different RNTIs. For example, DCI format 1A in the common search spaces may be used either with the C-RNTI for normal scheduling purposes or with the SI-RNTI for scheduling of system information. This does not affect the number of blind decoding attempts, as they are related to the DCI format; checking two different RNTIs, that is, checking two different CRCs, after decoding is a very low-complexity operation.

Copyright Elsevier 2024

Configuration of the transmission mode is done via RRC signaling. As the exact subframe number when this configuration takes effect in the terminal is not specified and may vary depending on, for example, RLC retransmissions, there is a (short) period when the network and the terminal may have different understandings of which transmission mode is configured. Therefore, in order not to lose the possibility of communicating with the terminal, it is necessary to have at least one DCI format that is decoded irrespective of the transmission mode. For downlink transmissions, DCI format 1A serves this purpose and the network can therefore always transmit data to the terminal using this DCI format. Another function of format 1A is to reduce overhead for transmissions when full flexibility in resource-block assignment is not needed.

10.4.6 Downlink scheduling assignments

Having described the transmission of downlink control information on PDCCH and EPDCCH, the detailed contents of the control information can be discussed, starting with the downlink scheduling assignments. Downlink scheduling assignments are valid for the same subframe in which they are transmitted. The scheduling assignments use one of the DCI formats 1, 1A, 1B, 1C, 1D, 2, 2A, 2B, 2C, or 2D and the DCI formats used depend on the transmission mode configured (see Table 10.5 for the relation between DCI formats and transmission modes). The reason for supporting multiple formats with different message sizes for the same purpose is to allow for a trade-off in control-signaling overhead and scheduling flexibility. Parts of the contents are the same for the different DCI formats, as seen in Table 10.6, but there are also differences due to the different capabilities.

DCI format 1 is the basic downlink assignment format in the absence of spatial multiplexing (transmission modes 1, 2, and 7). It supports non-contiguous allocations of resource blocks and the full range of modulation-and-coding schemes.

DCI format 1A, also known as the "compact" downlink assignment, supports allocation of frequency-contiguous resource blocks only and can be used in all transmission modes. Contiguous allocations reduce the payload size of the control information with a somewhat reduced flexibility in resource allocations. The full range of modulation-and-coding schemes is supported. Format 1A can also be used to trigger a contention-free random access (see Chapter 14), in which case some bit fields are used to convey the necessary random-access preamble information and the remaining bit fields are set to a specific combination.

DCI format 1B is used to support codebook-based beam-forming described in Section 10.3.2, with a low control-signaling overhead (transmission mode 6). The content is similar to DCI format 1A with the addition of bits for signaling of the precoding matrix. As codebook-based beam-forming can be used to improve the data rates for cell-edge terminals, it is important to keep the related DCI message size small so as not to unnecessarily limit the coverage.

DCI format 1C is used for various special purposes such as random-access response, paging, transmission of system information, and MBMS-related signaling (see Chapter 17).

Copyright Elsevier 2024

Samsung Exhibit 1017, Page 249 of 538
Samsung Electronics Co., Ltd. v. Asus Technology Licensing Inc.
IPR2024-01217

Table 10.6 DCI Formats Used for Downlink Scheduling

Field		DCI Format									
		1	**1A**	**1B**	**1C**	**1D**	**2**	**2A**	**2B**	**2C**	**2D**
Resource information	Carrier indicator	•	•	•		•	•	•	•	•	•
	Resource block assignment type	0/1	2	2	2'	2	0/1	0/1	0/1	0/1	0/1
HARQ process	number	•	•	•		•	•	•	•	•	•
1st transport block	MCS	•	•	•	•	•	•	•	•	•	•
	RV	•	•	•		•	•	•	•	•	•
	NDI	•	•	•		•	•	•	•	•	•
2nd transport block	MCS						•	•	•	•	•
	RV						•	•	•	•	•
	NDI						•	•	•	•	•
Multi-antenna information	PMI confirmation			•							
	Precoding information			•		•	•	•			
	Transport block swap flag						•	•			
	Power offset					•					
	DM-RS scrambling								•		
	#Layers/DM-RS scrambling/ antenna ports									•	•
PDSCH mapping and quasi-co-location indicator											•
Downlink assignment index		•	•	•		•	•	•	•	•	•
PUCCH power control		•	•	•		•	•	•	•	•	•
SRS request[a]			F						T	T	T
ACK/NAK offset (EPDCCH only)			•	•		•	•	•	•	•	•
Flag for 0/1A differentiation			•								
Padding (only if needed)		(•)	(•)	(•)		(•)	(•)	(•)	(•)	(•)	(•)
Identity		•	•	•	•	•	•	•	•	•	

[a]*Format 1A for FDD and formats 2B, 2C, and 2D for TDD.*

Common for these applications is simultaneous reception of a relatively small amount of information by *multiple* users. Hence, DCI format 1C supports QPSK only, has no support for hybrid-ARQ retransmissions, and does not support closed-loop spatial multiplexing. Consequently, the message size for DCI format 1C is very small, which is beneficial for

Copyright Elsevier 2024

coverage and efficient transmission of the type of system messages for which it is intended. Furthermore, as only a small number of resource blocks can be indicated, the size of the corresponding indication field in DCI format 1C is independent of the cell bandwidth.

DCI format 1D is used to support multi-user MIMO (transmission mode 5) scheduling of one codeword with precoder information. To support dynamic sharing of the transmission power between the terminals sharing the same resource block in multi-user MIMO, one bit of power offset information is included in DCI format 1D, as described in Section 10.3.4.2.

DCI format 2 is an extension for DCI format 1 to support closed-loop spatial multiplexing (transmission mode 4). Thus, information about the number of transmission layers and the index of the precoder matrix used are jointly encoded in the precoding information field. Some of the fields in DCI format 1 have been duplicated to handle the two transport blocks transmitted in parallel in the case of spatial multiplexing.

DCI format 2A is similar to DCI format 2 except that it supports open-loop spatial multiplexing (transmission mode 3) instead of closed-loop spatial multiplexing. The precoder information field is used to signal the number of transmission layers only, hence the field has a smaller size than in DCI format 2. Furthermore, since DCI format 2A is used for scheduling of multi-layer transmissions only, the precoder information field is only necessary in the case of four transmit antenna ports; for two transmit antennas the number of layers is implicitly given by the number of transport blocks.

DCI format 2B was introduced in release 9 in order to support dual-layer spatial multiplexing in combination with beam-forming using DM-RS (transmission mode 8). Since scheduling with DCI format 2B relies on DM-RS, precoding/beam-forming is transparent to the terminal and there is no need to signal a precoder index. The number of layers can be controlled by disabling one of the transport blocks. Two different scrambling sequences for the DM-RS can be used, as described in Section 10.2.2.

DCI format 2C was introduced in release 10 and is used to support spatial multiplexing using DM-RS (transmission mode 9). To some extent, it can be seen as a generalization of format 2B to support spatial multiplexing of up to eight layers. DM-RS scrambling and the number of layers are jointly signaled by a single three-bit field.

DCI format 2D was introduced in release 11 and is used to support spatial multiplexing using DM-RS (transmission mode 10). In essence, it is an extension of format 2C to support signaling of quasi-co-location of antenna ports.

Many information fields in the different DCI formats are, as already mentioned, common among several of the formats, while some types of information exist only in certain formats. Furthermore, in later releases, some of the DCI formats are extended with additional bits. One example is the addition of the carrier indicator in many DCI formats in release 10, as well as the inclusion of the SRS request in DCI formats 0 and 1A in release 10. Such extensions, for which different releases have different payload sizes for the same DCI format, are possible as long the extensions are used in the terminal-specific search spaces only where a single terminal is addressed by the DCI format in question. For DCI formats used to address multiple

Copyright Elsevier 2024

terminals at the same time, for example, to broadcast system information, payload extensions are clearly not possible as terminals from previous releases are not capable of decoding these extended formats. Consequently, the extensions are allowed in the terminal-specific search spaces only.

The information in the DCI formats used for downlink scheduling can be organized into different groups, as shown in Table 10.6, with the fields present varying between the DCI formats. The contents of the different DCI formats are explained in more detail below:

- Resource information, consisting of:
  - Carrier indicator (0 or 3 bit). This field is present in releases 10 and beyond only if cross-carrier scheduling is enabled via RRC signaling and is used to indicate the component carrier the downlink control information relates to (see Section 10.4.8). The carrier indicator is not present in the common search space as this would either impact compatibility with terminals not capable of carrier aggregation or require additional blind decoding attempts.
  - Resource-block allocation. This field indicates the resource blocks on one component carrier upon which the terminal should receive the PDSCH. The size of the field depends on the cell bandwidth and on the DCI format, more specifically on the resource indication type, as discussed in Section 10.4.6.1. Resource allocation types 0 and 1, which are the same size, support non-contiguous resource-block allocations, while resource allocation type 2 has a smaller size but supports contiguous allocations only. DCI format 1C uses a restricted version of type 2 in order to further reduce control signaling overhead.
  - Hybrid-ARQ process number (3 bit for FDD, 4 bit for TDD), informing the terminal about the hybrid-ARQ process to use for soft combining. Not present in DCI format 1C as this DCI format is intended for scheduling of system information which does not use hybrid-ARQ retransmissions.
- For the first (or only) transport block:[38]
  - Modulation-and-coding scheme (5 bit), used to provide the terminal with information about the modulation scheme, the code rate, and the transport-block size, as described further below. DCI format 1C has a restricted size of this field as only QPSK is supported.
  - New-data indicator (1 bit), used to clear the soft buffer for initial transmissions. Not present in DCI format 1C as this format does not support hybrid ARQ.
  - Redundancy version (2 bit).
- For the second transport block (only present in DCI formats supporting spatial multiplexing):
  - Modulation-and-coding scheme (5 bit).

[38]A transport block can be disabled by setting the modulation-and-coding scheme to zero and the redundancy version to one in the DCI.

Copyright Elsevier 2024

  - New-data indicator (1 bit).
  - Redundancy version (2 bit).
- Multi-antenna information. The different DCI formats are intended for different multi-antenna schemes and which of the fields below that are included depends on the DCI format shown in Table 10.5.
  - PMI confirmation (1 bit), present in format 1B only. Indicates whether the eNodeB uses the (frequency-selective) precoding matrix recommendation from the terminal or if the recommendation is overridden by the information in the PMI field.
  - Precoder-matrix information (PMI), provides information about the index of the precoding matrix used for the downlink transmission and, indirectly, about the number of transmission layers. This information is present in the DCI formats used for CRS-based transmission only; for DM-RS-based transmissions the precoder used is transparent for the terminal, and consequently there is no need to signal this information in this case.
  - Transport block swap flag (1 bit), indicating whether the two codewords should be swapped prior to being fed to the hybrid-ARQ processes. Used for averaging the channel quality between the codewords.
  - Power offset between the PDSCH and cell-specific reference signals used to support dynamic power sharing between multiple terminals for multi-user MIMO.
  - Reference-signal scrambling sequence, used to control the generation of quasi-orthogonal DM-RS sequences, as discussed in Section 10.2.2.
  - Number of layers, reference-signal scrambling sequence, and the set of antenna ports used for the transmission (jointly encoded information in releases 10 and later).
- PDSCH resource-element mapping and quasi-co-location indicator (2 bit), informing the terminal which set of parameters to assume when demodulating the PDSCH. Up to four different sets of parameters can be configured by RRC to support different CoMP schemes, as discussed in Chapter 15.
- Downlink assignment index (2 bit), informing the terminal about the number of downlink transmissions for which a single hybrid-ARQ acknowledgement should be generated according to Section 12.1.3. Present for TDD only.
- Transmit-power control for PUCCH (2 bit). For scheduling of a secondary component carrier in the case of carrier aggregation, these bits are reused as *acknowledgement resource indicator* (ARI)—see Chapter 11.
- SRS request (1 bit). This field is used to trigger a one-shot transmission of a sounding reference signal in the uplink, a feature introduced in release 10 and an example of extending an existing DCI format in the terminal-specific search space with additional information fields in a later release. For FDD it is present in format 1A only, while for TDD it is present in formats 2B, 2C, and 2D only. For TDD, where short-term channel reciprocity can be used, it is motivated to include in this field the DCI formats used to

Copyright Elsevier 2024

support DM-RS-based multi-layer transmission in order to estimate the downlink channel conditions based on uplink channel sounding. For FDD, on the other hand, this would not be useful, and the SRS request field is consequently not included. However, for DCI format 1A, including a SRS request bit comes "for free," as padding otherwise would have to be used to ensure the same payload size as for DCI format 0.

- ACK/NAK offset (2 bit). This field is present on EPDCCH only and thus is supported in release 11 only. It is used to dynamically control the PUCCH resource used for the hybrid-ARQ acknowledgement as discussed in Section 11.4.1.
- DCI format 0/1A indication (1 bit), used to differentiate between DCI formats 1A and 0 as the two formats have the same message size. This field is present in DCI formats 0 and 1A only. DCI formats 3 and 3A, which have the same size, are separated from DCI formats 0 and 1A through the use of a different RNTI.
- Padding. The smaller of DCI formats 0 and 1A is padded to ensure the same payload size irrespective of the uplink and downlink cell bandwidths. Padding is also used to ensure that the DCI size is different for different DCI formats that may occur simultaneously in the same search space (this is rarely required in practice as the payload sizes are different due to the different amounts of information). Finally, for PDCCH, padding is used to avoid certain DCI sizes that may cause ambiguous decoding.[39]
- Identity (RNTI) of the terminal for which the PDSCH transmission is intended (16 bit). As described in Sections 10.4.3 and 10.4.4, the identity is not explicitly transmitted but implicitly included in the CRC calculation. There are different RNTIs defined depending on the type of transmission (unicast data transmission, paging, power-control commands, etc.).

10.4.6.1 Signaling of downlink resource-block allocations

Focusing on the signaling of resource-block allocations, there are three different possibilities, types 0, 1, and 2, as indicated in Table 10.6. Resource-block allocation types 0 and 1 both support non-contiguous allocations of resource blocks in the frequency domain, whereas type 2 supports contiguous allocations only. A natural question is why multiple ways of signaling the resource-block allocations are supported, a question whose answer lies in the number of bits required for the signaling. The most flexible way of indicating the set of resource blocks the terminal is supposed to receive the downlink transmission upon is to include a bitmap with size equal to the number of resource blocks in the cell bandwidth. This would allow for an arbitrary combination of resource blocks to be scheduled for transmission to the terminal but

[39]For a small set of specific payload sizes, the control signaling on PDCCH may be correctly decoded at an aggregation level other than the one used by the transmitter. Since the PHICH resource is derived from the first CCE used for the PDCCH, this may result in incorrect PHICH being monitored by the terminal. To overcome this, padding is used if necessary to avoid the problematic payload sizes. Note that this padding applies for PDCCH only.

Copyright Elsevier 2024

would, unfortunately, also result in a very large bitmap for the larger cell bandwidths. For example, in the case of a downlink cell bandwidth corresponding to 100 resource blocks, the downlink PDCCH would require 100 bits for the bitmap alone, to which the other pieces of information need to be added. Not only would this result in a large control-signaling overhead, but it could also result in downlink coverage problems as more than 100 bits in one OFDM symbol correspond to a data rate exceeding 1.4 Mbit/s. Consequently, there is a need for a resource allocation scheme requiring a smaller number of bits while keeping sufficient allocation flexibility.

In resource allocation type 0, the size of the bitmap has been reduced by pointing not to individual resource blocks in the frequency domain, but to groups of contiguous resource blocks, as shown at the top of Figure 10.34. The size of such a group is determined by the downlink cell bandwidth. For the smallest bandwidths there is only a single resource block in a group, implying that an arbitrary set of resource blocks can be scheduled, whereas for the largest cell bandwidths, groups of four resource blocks are used (in the example in Figure 10.34, the cell bandwidth is 25 resource blocks, implying a group size of two resource blocks). Thus, the bitmap for the system with a downlink cell bandwidth of 100 resource blocks is reduced from 100 to 25 bits. A drawback is that the scheduling granularity is reduced; single resource blocks cannot be scheduled for the largest cell bandwidths using allocation type 0.

However, also in large cell bandwidths, frequency resolution of a single resource block is sometimes useful, for example to support small payloads. Resource allocation type 1 addresses this by dividing the total number of resource blocks in the frequency domain into dispersed

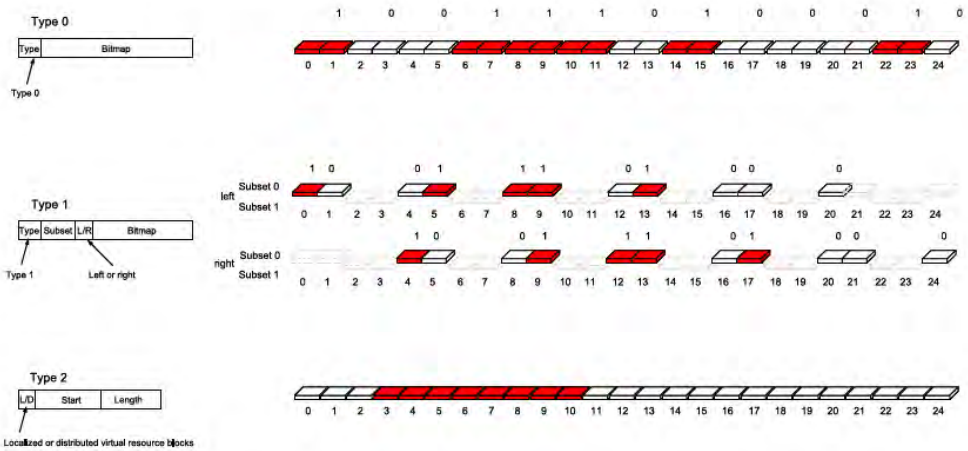


FIGURE 10.34

Illustration of resource-block allocation types (cell bandwidth corresponding to 25 resource blocks used in this example)

Copyright Elsevier 2024

subsets, as shown in the middle of Figure 10.34. The number of subsets is given from the cell bandwidth with the number of subsets in type 1 being equal to the group size in type 0. Thus, in Figure 10.34, there are two subsets, whereas for a cell bandwidth of 100 resource blocks there would have been four different subsets. Within a subset, a bitmap indicates the resource blocks in the frequency domain upon which the downlink transmission occurs.

To inform the terminal whether resource allocation type 0 or 1 is used, the resource allocation field includes a flag for this purpose, denoted "type" to the left in Figure 10.34. For type 0, the only additional information is the bitmap discussed previously. For type 1, on the other hand, in addition to the bitmap itself, information about the subset for which the bitmap relates is also required. As one of the requirements in the design of resource allocation type 1 was to maintain the same number of bits in the allocation as for type 0 without adding unnecessary overhead,[40] the bitmap in resource allocation type 1 is smaller than in type 0 to allow for the signaling of the subset number. However, a consequence of a smaller bitmap is that not all resource blocks in the subset can be addressed simultaneously. To be able to address all resources with the bitmap, there is a flag indicating whether the bitmap relates to the "left" or "right" part of the resource blocks, as depicted in the middle part of Figure 10.34.

Unlike the other two types of resource-block allocation signaling, type 2 does not rely on a bitmap. Instead, it encodes the resource allocation as a start position and length of the resource-block allocation. Thus, it does not support arbitrary allocations of resource blocks but only frequency-contiguous allocations, thereby reducing the number of bits required for signaling the resource-block allocation. The number of bits required for resource-signaling type 2 compared to type 0 or 1 is shown in Figure 10.35; the difference is fairly large for the larger cell bandwidths.

All three resource-allocation types refer to *virtual* resource blocks (see Section 10.1.1.8 for a discussion of resource-block types). For resource-allocation types 0 and 1, the virtual resource blocks are of localized type and the virtual resource blocks are directly mapped to physical resource blocks. For resource-allocation type 2, on the other hand, both localized and distributed virtual resource blocks are supported. One bit in the resource allocation field indicates whether the allocation signaling refers to localized or distributed resource blocks.

10.4.6.2 Signaling of transport-block sizes

Proper reception of a downlink transmission requires, in addition to the set of resource blocks, knowledge about the modulation scheme and the transport-block size, information (indirectly) provided by a 5-bit field in the different DCI formats. Of the 32 combinations, 29 are used to signal the modulation-and-coding scheme whereas three are reserved, the purpose of which is described later. Together, the modulation-and-coding scheme and the number of resource blocks assigned provide the transport-block size on the DL-SCH. Thus, the possible transport-block sizes can be described as a table with 29 rows and 110 columns, one column

[40] Allowing different sizes would result in an increase in the number of blind decoding attempts required in the terminal.

Copyright Elsevier 2024

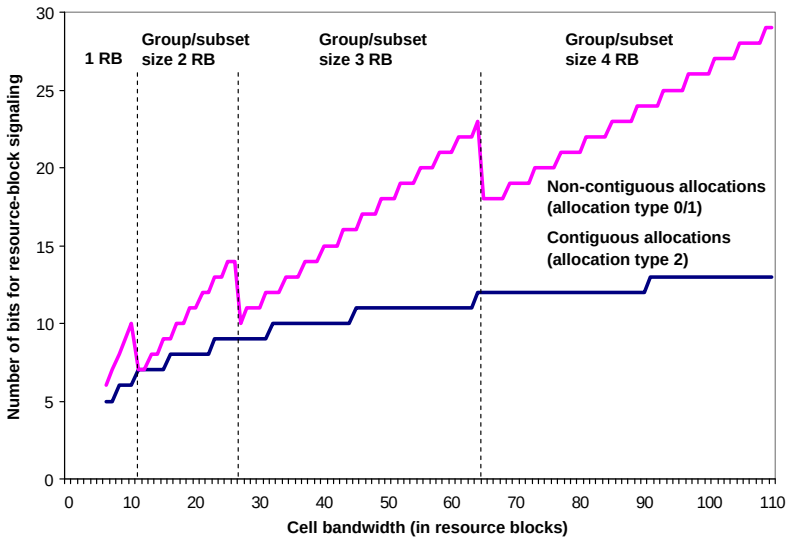


FIGURE 10.35

Number of bits used for downlink resource-allocation signaling for downlink allocation types 0/1 and 2

for each number of resource blocks possible to transmit upon (the number of columns follows from the maximum downlink component carrier bandwidth of 110 resource blocks).

Each modulation-and-coding scheme represents a particular combination of modulation scheme and channel-coding rate or, equivalently, a certain spectral efficiency measured in the number of information bits per modulation symbol. Although the 29-by-110 table of transport-block sizes in principle could be filled directly from the modulation-and-coding scheme and the number of resource blocks, this would result in arbitrary transport-block sizes, which is not desirable. First, as all the higher-layer protocol layers are byte aligned, the resulting transport-block sizes should be an integer number of bytes. Secondly, common payloads, for example RRC signaling messages and VoIP, should be possible to transmit without padding. Aligning with the QPP interleaver sizes is also beneficial, as this would avoid the use of filler bits (see Section 10.1.1.3). Finally, the same transport-block size should ideally appear for several different resource-block allocations, as this allows the number of resource blocks to be changed between retransmission attempts, providing increased scheduling flexibility. Therefore, a "mother table" of transport-block sizes is first defined, fulfilling the requirements above. Each entry in the 29-by-110 table is picked from the mother table such that the resulting spectral efficiency is as close as possible to the spectral efficiency of the signaled modulation-and-coding scheme. The mother table spans the full range of transport-block sizes possible, with an approximately constant worst-case padding.

Copyright Elsevier 2024

From a simplicity perspective, it is desirable if the transport-block sizes do not vary with the configuration of the system. The set of transport-block sizes is therefore independent of the actual number of antenna ports and the size of the control region.[41] The design of the table assumes a control region of three OFDM symbols and two antenna ports, the "reference configuration." If the actual configuration is different, the resulting code rate for the DL-SCH will be slightly different as a result of the rate-matching procedure. However, the difference is small and of no practical concern. Also, if the actual size of the control region is smaller than the three-symbol assumption in the reference configuration, the spectral efficiencies will be somewhat smaller than the range indicated by the modulation-and-coding scheme signaled as part of the DCI. Thus, information about the modulation scheme used is obtained directly from the modulation-and-coding scheme, whereas the exact code rate and rate matching is obtained from the implicitly signaled transport-block size together with the number of resource elements used for DL-SCH transmission.

For bandwidths smaller than the maximum of 110 resource blocks, a subset of the table is used. More specifically, in case of a cell bandwidth of N resource blocks, the first N columns of the table are used. Also, in the case of spatial multiplexing, a single transport block can be mapped to up to four layers. To support the higher data rates this facilitates, the set of supported transport-block sizes needs to be extended beyond what is possible in the absence of spatial multiplexing. The additional entries are in principle obtained by multiplying the sizes with the number of layers to which a transport block is mapped and adjusting the result to match the QPP interleaver size.

The 29 combinations of modulation-and-coding schemes each represent a reference spectral efficiency in the approximate range of 0.1–5.9 bits/s/symbol.[42] There is some overlap in the combinations in the sense that some of the 29 combinations represent the same spectral efficiency. The reason is that the best combination for realizing a specific spectral efficiency depends on the channel properties; sometimes higher-order modulation with a low code rate is preferable over lower-order modulation with a higher code rate, and sometimes the opposite is true. With the overlap, the eNodeB can select the best combination, given the propagation scenario. As a consequence of the overlap, two of the rows in the 29-by-110 table are duplicates and result in the same spectral efficiency but with different modulation schemes, and there are only 27 unique rows of transport-block sizes.

Returning to the three reserved combinations in the modulation-and-coding field mentioned at the beginning of this section, those entries can be used for retransmissions only. In the case of a retransmission, the transport-block size is, by definition, unchanged and fundamentally there is no need to signal this piece of information. Instead, the three reserved values represent the modulation scheme, QPSK, or 16QAM or 64QAM, which allows the scheduler to use an (almost) arbitrary combination of resource blocks for the retransmission.

[41]For DwPTS, the transport-block size is scaled by a factor of 0.75 compared to the values found in the table, motivated by the DwPTS having a shorter duration than a normal subframe.
[42]The exact values vary slightly with the number of resource blocks allocated due to rounding.

Copyright Elsevier 2024

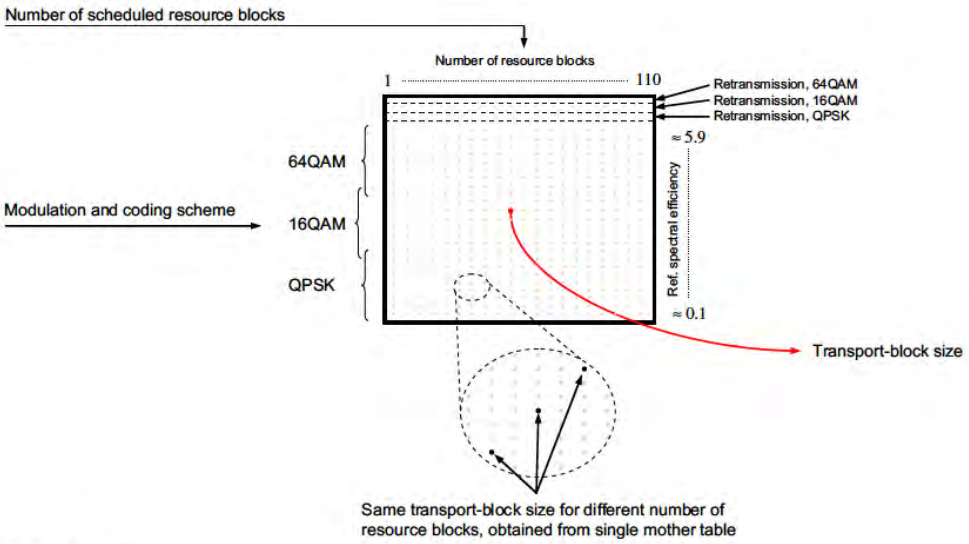


FIGURE 10.36

Computing the transport-block size

Using any of the three reserved combinations assumes that the terminal properly received the control signaling for the initial transmission; if this is not the case, the retransmission should explicitly indicate the transport-block size.

The derivation of the transport-block size from the modulation-and-coding scheme and the number of scheduled resource blocks is illustrated in Figure 10.36.

10.4.7 Uplink scheduling grants

Uplink scheduling grants use one of DCI formats 0 or 4; DCI format 4 was added in release 10 to support uplink spatial multiplexing. The basic resource-allocation scheme for the uplink is single-cluster allocations where the resource blocks are contiguous in the frequency domain, although release 10 added support for multi-cluster transmissions of up to two clusters on a single component carrier.

DCI format 0 is used for scheduling uplink transmissions not using spatial multiplexing on one component carrier. It has the same size control-signaling message as the "compact" downlink assignment (DCI format 1A). A flag in the message is used to inform the terminal whether the message is an uplink scheduling grant (DCI format 0) or a downlink scheduling assignment (DCI format 1A).

Copyright Elsevier 2024

Table 10.7 DCI Formats for Uplink Scheduling Grants

Field		DCI Format	
		0	4
Resource information	Carrier indicator	•	•
	Resource allocation type	•	•
	Resource block assignment	0/(1)	0/1
1st transport block	MCS/RV	•	•
	NDI	•	•
2nd transport block	MCS/RV		•
	NDI		•
DM-RS phase rotation and OCC index		•	•
Precoding information			•
CSI request		•	•
SRS request		•	•
Uplink index/DAI (TDD only)		•	•
PUSCH power control		•	•
Flag for 0/1A differentiation		•	
Padding (only if needed)		(•)	(•)
Identity		•	•

DCI format 4 is used for uplink transmissions using spatial multiplexing on one component carrier. Consequently, the size of DCI format 4 is larger than that of DCI format 0, as additional information fields are required.

Many information fields are common to the two DCI formats, but there are also differences, as shown in Table 10.7. Similarly to the downlink scheduling assignments, some DCI formats have been extended with additional bits in later releases. The contents of the different DCI formats are explained in more detail below:

- Resource information, consisting of:
  - Carrier indicator (0 or 3 bit). This field is present in releases 10 and beyond only if cross-carrier scheduling is enabled via RRC signaling and is used to indicate the uplink component carrier the grant relates to (see Section 10.4.8). The carrier indicator is not present in the common search space as this would either impact compatibility with terminals not capable of carrier aggregation or require additional blind decoding attempts.
  - Resource allocation type or multi-cluster flag (1 bit), indicating whether resource allocation type 0 (one cluster of resource blocks is used for uplink transmissions) or type 1 (two clusters of resource blocks are used for the uplink transmission) is used.

Copyright Elsevier 2024

This flag is not present in releases prior to release 10. In previous releases, the downlink bandwidth was, in practice, always at least as large as the uplink bandwidth, implying that one padding bit was used for DCI format 0 in those releases to align with the size of format 1A. The padding bit could therefore be replaced by the multi-cluster flag in release 10 without sacrificing backwards compatibility. In DCI format 4, supported in releases 10 and later, the multi-cluster flag is always present.
    - Resource-block allocation, including hopping indication. This field indicates the resource blocks upon which the terminal should transmit the PUSCH using uplink resource-allocation type 0 (DCI format 0) or type 1 (DCI format 4), as described in Section 10.4.7.1. The size of the field depends on the cell bandwidth. For single-cluster allocations, uplink frequency hopping, as described in Chapter 11, can be applied to the uplink PUSCH transmission.
- For the first (or only) transport block:
    - Modulation-and-coding scheme including redundancy version (5 bit), used to provide the terminal with information about the modulation scheme, the code rate, and the transport-block size. The signaling of the transport-block size uses the same transport-block table as for the downlink—that is, the modulation-and-coding scheme together with the number of scheduled resource blocks provides the transport-block size. However, as the support of 64QAM in the uplink is not mandatory for all terminals, terminals not capable of 64QAM use 16QAM when 64QAM is indicated in the modulation-and-coding field. The use of the three reserved combinations is slightly different than for the downlink; the three reserved values are used for implicit signaling of the redundancy version, as described below. A transport block can be disabled by signaling a specific combination of modulation-and-coding scheme and number of resource blocks. Disabling one transport block is used when retransmitting a single transport block only.
    - New-data indicator (1 bit), used to indicate to the terminal whether transmission of a new transport block or retransmission of the previous transport block is granted.
- For the second transport block (DCI format 4 only):
    - Modulation-and-coding scheme including redundancy version (5 bit).
    - New-data indicator (1 bit).
- Phase rotation of the uplink demodulation reference signal (3 bit), used to support multi-user MIMO, as described in Chapter 11. By assigning different reference-signal phase rotations to terminals scheduled on the same time–frequency resources, the eNodeB can estimate the uplink channel response from each terminal and suppress the inter-terminal interference by the appropriate processing. In releases 10 and later, it also controls the orthogonal cover sequence (see Section 11.2).
- Precoding information, used to signal the precoder to use for the uplink transmission in releases 10 and later.
- Channel-state request flag (1 or 2 bit). The network can explicitly request an aperiodic channel-state report to be transmitted on the UL-SCH by setting this bit(s) in the

Copyright Elsevier 2024

uplink grant. In the case of carrier aggregation, two bits are used to indicate which downlink component carrier the CSI should be reported for (see Chapter 12).

- SRS request (2 bit), used to trigger aperiodic sounding using one of up to three preconfigured settings, as discussed in Chapter 11. The SRS request, introduced in release 10, is supported in the terminal-specific search space only for reasons already described.
- Uplink index/DAI (2 bit). This field is present only when operating in TDD. For uplink–downlink configuration 0 (uplink-heavy configuration), it is used as an uplink index to signal for which uplink subframe(s) the grant is valid, as described in Chapter 13. For other uplink–downlink configurations, it is used as a downlink assignment index to indicate the number of downlink transmissions the eNodeB expects hybrid-ARQ acknowledgement for.
- Transmit-power control for PUSCH (2 bit).
- DCI format 0/1A indication (1 bit), used to differentiate between DCI formats 1A and 0 as the two formats have the same message size. This field is present in DCI formats 0 and 1A only.
- Padding. The smaller of DCI formats 0 and 1A is padded to ensure the same payload size irrespective of the uplink and downlink cell bandwidths. Padding is also used to ensure that the DCI size is different for DCI formats 0 and 4 (this is rarely required in practice as the payload sizes are different due to the different amounts of information). Finally, for PDCCH, padding is used to avoid certain DCI sizes that may cause ambiguous decoding.
- Identity (RNTI) of the terminal for which the grant is intended (16 bit). As described in Sections 10.4.3 and 10.4.4, the identity is not explicitly transmitted but implicitly included in the CRC calculation.

There is no explicit signaling of the redundancy version in the uplink scheduling grants. This is motivated by the use of a synchronous hybrid-ARQ protocol in the uplink; retransmissions are normally triggered by a negative acknowledgement on the PHICH and not explicitly scheduled as for downlink data transmissions. However, as described in Chapter 12, there is a possibility to explicitly schedule retransmissions. This is useful in a situation where the network will explicitly move the retransmission in the frequency domain by using the PDCCH instead of the PHICH. Three values of the modulation-and-coding field are reserved to mean redundancy versions one, two, and three. If one of those values is signaled, the terminal should assume that the same modulation and coding as the original transmission is used. The remaining entries are used to signal the modulation-and-coding scheme to use and also imply that redundancy version zero should be used. The difference in usage of the reserved values compared to the downlink scheduling assignments implies that the modulation scheme, unlike the downlink case, cannot change between uplink (re)transmission attempts.

Copyright Elsevier 2024

10.4.7.1 Signaling of uplink resource-block allocations

The basic uplink resource-allocation scheme is single-cluster allocations—that is, allocations contiguous in the frequency domain—but releases 10 and later also provide the possibility for multi-cluster uplink transmissions.

Single-cluster allocations use uplink resource-allocation type 0, which is identical to downlink resource allocation type 2 described in Section 10.4.6.1 except that the single-bit flag indicating localized/distributed transmission is replaced by a single-bit frequency hopping flag. The resource allocation field in the DCI provides the set of virtual resource blocks to use for uplink transmission. The set of physical resource blocks to use in the two slots of a subframe is controlled by the hopping flag, as described in Chapter 11.

Multi-cluster allocations with up to two clusters were introduced in release 10, using uplink resource-allocation type 1. In resource-allocation type 1, the starting and ending positions of two clusters of frequency-contiguous resource blocks are encoded into an index. Uplink resource-allocation type 1 does not support frequency hopping (diversity is achieved through the use of two clusters instead). Indicating two clusters of resources naturally requires additional bits compared to the single-cluster case. At the same time, the total number of bits used for resource-allocation type 1 should be identical to that of type 0. This is similar to the situation for the downlink allocation types 0 and 1; without aligning the sizes a new DCI format with a corresponding negative impact on the number of blind decodings is necessary. Since frequency hopping is not supported for allocation type 1, the bit otherwise used for the hopping flag can be reused for extending the resource-allocation field. However, despite the extension of the resource-allocation field by one bit, the number of bits is not sufficient to provide a single-resource-block resolution in the two clusters for all bandwidths. Instead, similar to downlink resource-allocation type 0, groups of resource blocks are used and the starting and ending positions of the two clusters are given in terms of group numbers. The size of such a group is determined by the uplink carrier bandwidth in a similar way as for the downlink. For the smallest bandwidths there is only a single resource block in a group, implying that an arbitrary (as long as the limit of at most two clusters is observed) set of resource blocks can be scheduled, whereas for the largest cell bandwidths groups of four resource blocks are used. In the example in Figure 10.37, the cell bandwidth is 25 resource blocks, implying a group size of two resource blocks.

10.4.8 Carrier aggregation and cross-carrier scheduling

Carrier aggregation, where a terminal receives or transmits on multiple component carriers, is an integral part of LTE from releases 10 onwards. For a terminal supporting multiple component carriers, the terminal needs to know to which component carrier a certain DCI relates. This information can either be implicit or explicit, depending on whether cross-carrier scheduling is used or not, as illustrated in Figure 10.38. Enabling cross-carrier scheduling is done individually via RRC signaling on a per-terminal and per-component-carrier basis.

Copyright Elsevier 2024

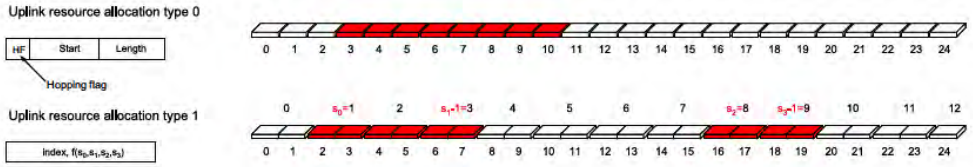


FIGURE 10.37

Illustration of uplink resource-block allocation types (uplink bandwidth corresponding to 25 resource blocks used in this example)

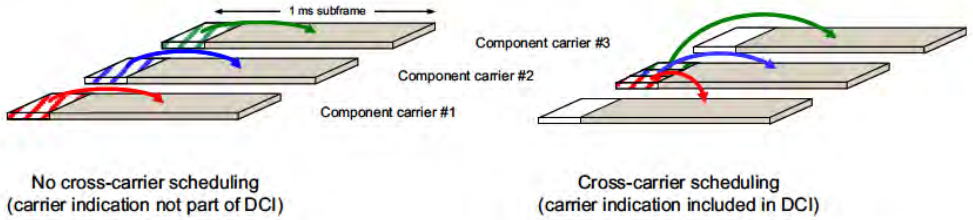


FIGURE 10.38

Cross-carrier scheduling

In the absence of cross-carrier scheduling, downlink scheduling assignments are valid for the component carrier upon which they are transmitted. Similarly, for uplink grants, there is an association between downlink and uplink component carriers such that each uplink component carrier has an associated downlink component carrier. The association is provided as part of the system information. Thus, from the uplink–downlink association, the terminal will know to which uplink component carrier the downlink control information relates to.

In the presence of cross-carrier scheduling, where downlink PDSCH or uplink PUSCH is transmitted on an (associated) component carrier other than that which (E)PDCCH is transmitted upon, the carrier indicator in the PDCCH provides information about the component carrier used for the PDSCH or PUSCH.

Whether cross-carrier scheduling is used or not is, as already mentioned, configured using higher-layer signaling; if cross-carrier scheduling is not configured then no carrier indication field is included in the DCI. Thus, most of the DCI formats come in two "flavors," with and without the carrier indication field, and which "flavor" the terminal is supposed to monitor is determined by enabling/disabling support for cross-carrier scheduling.

To signal which component carrier a grant relates to, the component carriers are numbered. The primary component carrier is always given the number zero, while the different secondary component carriers are assigned a unique number each through terminal-specific RRC signaling. Hence, even if the terminal and the eNodeB may have different

Copyright Elsevier 2024

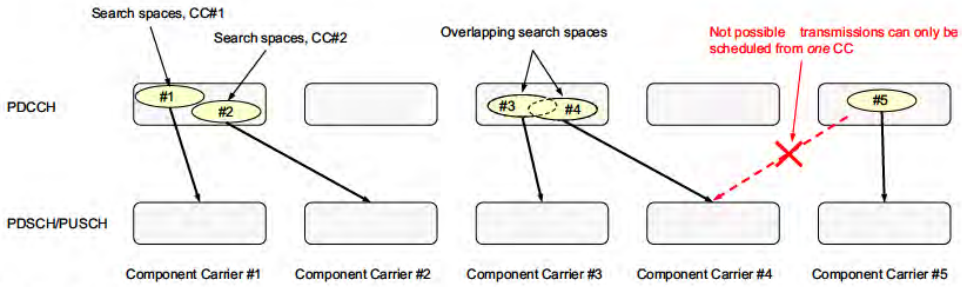


FIGURE 10.39

Example of scheduling of multiple component carriers (search spaces are discussed in Section 10.4.5)

understandings of the component carrier numbering during a brief period of reconfiguration, at least transmissions on the primary component carrier can be scheduled.

Irrespective of whether cross-carrier scheduling is used or not, PDSCH/PUSCH on a component carrier can only be scheduled from *one* component. Thus, for each PDSCH/PUSCH component carrier there is an associated component carrier, configured via terminal-specific RRC signaling, where the corresponding DCI can be transmitted. Figure 10.39 illustrates one example, where PDSCH/PUSCH transmissions on component carrier 1 are scheduled using PDCCHs/EPDCCHs transmitted on component carrier 1. In this case, as cross-carrier scheduling is not used, there is no carrier indicator in the corresponding DCI formats. PDSCH/PUSCH transmissions on component carrier 2 are cross-carrier scheduled from PDCCHs transmitted on component carrier 1. Hence, the DCI formats in the terminal-specific search space for component carrier 2 include the carrier indicator.

Note also, as transmissions on a component carrier can be scheduled by PDCCHs/EPDCCHs on *one* component carrier only, component carrier 4 cannot be scheduled by PDCCHs on component carrier 5 as the semi-static association between the component carriers used for PDCCH/EPDCCH transmission and the actual data transmission has associated data on component carrier 4, with PDCCHs/EPDCCHs on component carrier 3 in this example.

With respect to blind decoding and search spaces, the procedures described in Section 10.4.5 apply to each of the activated[43] downlink component carriers. Hence, in principle, there is one terminal-specific search space per aggregation level and per (activated) component carrier upon which PDSCH can be received (or PUSCH transmitted), although there are some carrier-aggregation-specific modifications. For terminals configured to use carrier aggregation, this results in an increase in the number of blind decoding attempts compared to a terminal not using carrier aggregation, as scheduling assignments/grants for each of the component carriers need to be monitored.

[43]Individual component carriers can be activated/deactivated as discussed in Chapter 12.

Copyright Elsevier 2024

The common search space is only defined for transmissions on the primary component carrier. As the main function of the common search space is to handle scheduling of system information intended for multiple terminals, and such information must be receivable by all terminals in the cell, scheduling in this case uses the common search space. For this reason, the carrier indication field is never present in DCI formats monitored in the common search space.

As mentioned before, there is one terminal-specific search space per aggregation level and component carrier[44] used for scheduling the PDSCH/PUSCH. This is illustrated in Figure 10.39, where PDSCH/PUSCH transmissions on component carrier 1 are scheduled using PDCCHs transmitted on component carrier 1. No carrier indicator is assumed in the terminal-specific search space for component carrier 1, as cross-carrier scheduling is not used. For component carrier 2, on the other hand, a carrier indicator is assumed in the terminal-specific search space, as component carrier 2 is cross-carrier scheduled from PDCCHs transmitted on component carrier 1.

Search spaces for different component carriers may overlap in some subframes. In Figure 10.39, this happens for the terminal-specific search spaces for component carriers 3 and 4. The terminal will handle the two search spaces independently, assuming (in this example) a carrier indicator for component carrier 4, but not for component carrier 3. If the terminal-specific and common search spaces relating to different component carriers happen to overlap for some aggregation level when cross-carrier scheduling is configured, the terminal only needs to monitor the common search space. The reason for this is to avoid ambiguities; if the component carriers have different bandwidths, a DCI format in the common search space may have the same payload size as another DCI format in the terminal-specific search space relating to another component carrier.

10.4.9 Power-control commands

As a complement to the power-control commands provided as part of the downlink scheduling assignments and the uplink scheduling grants, there is the potential to transmit a power-control command using DCI formats 3 (two-bit command per terminal) or 3A (single-bit command per terminal). The main motivation for DCI format 3/3A is to support power control for semi-persistent scheduling. The power-control message is directed to a group of terminals using an RNTI specific for that group. Each terminal can be allocated two power-control RNTIs, one for PUCCH power control and the other for PUSCH power control. Although the power-control RNTIs are common to a group of terminals, each terminal is informed through RRC signaling which bit(s) in the DCI message it should follow. No carrier indicator is used for formats 3 and 3A.

[44]If the DCI formats for two different component carriers have the same size, then search spaces can be shared such that, in the example in Figure 10.38, component carrier 2 can be scheduled from search spaces 2 and 1. This comes '"for free"' as the DCI formats in search space 1 need to be monitored anyway and the carrier-indicator field points to the component carrier for which the grant/assignment is intended.

Copyright Elsevier 2024

CHAPTER

Uplink Physical-Layer Processing 11

This chapter provides a description of the basic physical-layer functionality related to the LTE uplink. It essentially follows the same outline as the corresponding downlink description provided in the previous chapter, with detailed descriptions regarding transport-channel processing (Section 11.1), the reference-signal structure (Section 11.2), multi-antenna transmission (Section 11.3) and the uplink L1/L2 control-channel structure (Section 11.4). The chapter ends with an overview of uplink power control and the uplink timing-alignment procedure in Section 11.5 and Section 11.6 respectively. Physical aspects related to some specific uplink functions and procedures such as random access will be provided in later chapters.

11.1 Transport-channel processing

This section describes the physical-layer processing applied to the Uplink Shared Channel (UL-SCH) and the subsequent mapping to the uplink physical resource in the form of the basic OFDM time–frequency grid. As mentioned before, the UL-SCH is the only uplink transport-channel type in LTE[1] and is used for transmission of all uplink higher-layer information—that is, for both user data and higher-layer control signaling.

11.1.1 Processing steps

Figure 11.1 outlines the different steps of the UL-SCH physical-layer processing in case of transmission on a single carrier. Similar to the downlink, in the case of uplink carrier aggregation the different component carriers correspond to separate transport channels with separate physical-layer processing.

The different steps of the uplink transport-channel processing are summarized below. Most of these steps are very similar to the corresponding steps for DL-SCH processing outlined in Section 10.1. For a more detailed overview of the different steps, the reader is referred to the corresponding downlink description.

- *CRC insertion per transport block.* A 24-bit CRC is calculated for and appended to each transport block.

[1]Strictly speaking, the LTE Random-Access Channel is also defined as a transport-channel type (see Chapter 8). However, RACH only includes a layer-1 preamble and carries no data in the form of transport blocks.

4G: LTE/LTE-Advanced for Mobile Broadband. http://dx.doi.org/10.1016/B978-0-12-419985-9.00011-8

Copyright © 2014 Erik Dahlman, Stefan Parkvall and Johan Sköld. Published by Elsevier Ltd. All rights reserved.

Copyright Elsevier 2024

Samsung Exhibit 1017, Page 267 of 538
Samsung Electronics Co., Ltd. v. Asus Technology Licensing Inc.
IPR2024-01217

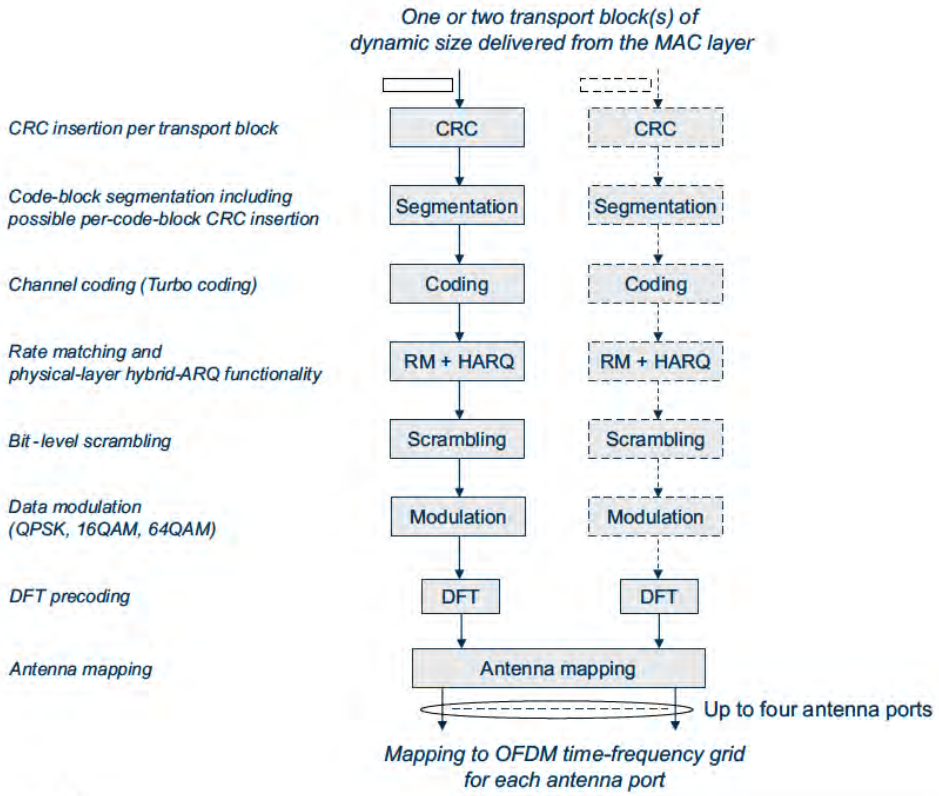


FIGURE 11.1

Physical-layer processing for the uplink shared channel (UL-SCH)

- *Code-block segmentation and per-code-block CRC insertion.* In the same way as for the downlink, code-block segmentation, including per-code-block CRC insertion, is applied for transport blocks larger than 6144 bits.
- *Channel coding.* Rate-1/3 Turbo coding with QPP-based inner interleaving is also used for the uplink shared channel.
- *Rate matching and physical-layer hybrid-ARQ functionality.* The physical-layer part of the uplink rate-matching and hybrid-ARQ functionality is essentially the same as the corresponding downlink functionality, with sub-block interleaving and insertion into a circular buffer followed by bit selection with four redundancy versions. There are some important differences between the downlink and uplink hybrid-ARQ protocols, such as

Copyright Elsevier 2024

asynchronous versus synchronous operation as described in Chapter 12. However, these differences are not visible in terms of the physical-layer processing.

- *Bit-level scrambling.* The aim of uplink scrambling is the same as for the downlink—that is, to randomize the, in this case, *uplink* interference to ensure that the processing gain provided by the channel code can be fully utilized.
- *Data modulation.* Similar to the downlink, QPSK, 16QAM, and 64QAM modulation can also be used for uplink shared-channel transmission.
- *DFT precoding.* As illustrated in Figure 11.2, the modulation symbols, in blocks of M symbols, are fed through a size-M DFT, where M corresponds to the number of subcarriers assigned for the transmission. The reason for the DFT precoding is to reduce the cubic metric for the transmitted signal. From an implementation complexity point-of-view the DFT size should preferably be constrained to a power of 2. However, such a constraint would limit the scheduler flexibility in terms of the amount of resources that can be assigned for an uplink transmission. Rather, from a flexibility point-of-view all possible DFT sizes should preferably be allowed. For LTE, a middle-way has been adopted where the DFT size, and thus also the size of the resource allocation, is limited to products of the integers 2, 3, and 5. For example, DFT sizes of 60, 72, and 96 are allowed but a DFT size of 84 is not allowed.[2] In this way, the DFT can be implemented as a combination of relatively low-complex radix-2, radix-3, and radix-5 FFT processing.
- *Antenna mapping.* The antenna mapping maps the output of the DFT precoder to one or several uplink antenna ports for subsequent mapping to the physical resource (the OFDM time–frequency grid). In the first releases of LTE (releases 8 and 9), only single-antenna transmission was used for the uplink.[3] However, as part of LTE release 10, support for uplink multi-antenna transmission by means of antenna precoding with up to four antennas ports was introduced. More details about LTE uplink multi-antenna transmission are provided in Section 11.3.

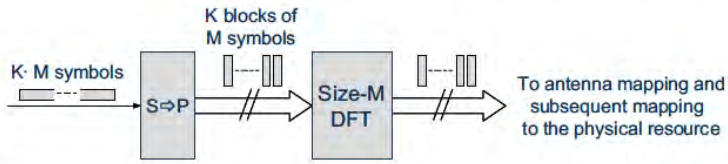


FIGURE 11.2

DFT precoding of K blocks, each consisting of M modulation symbols

[2]As uplink resource assignments are always done in terms of resource blocks of size 12 subcarriers, the DFT size is always a multiple of 12.

[3]Uplink multi-antenna transmission in the form of *antenna selection* has been part of the LTE specification since release 8. However, it is an optional terminal feature with very limited, if any, implementation in commercially available terminals.

Copyright Elsevier 2024

11.1.2 Mapping to the physical resource

The scheduler assigns a set of resource-block pairs to be used for the uplink transmission, more specifically for transmission of the PUSCH physical channel that carries the UL-SCH transport channel. Each such resource-block pair spans 14 OFDM symbols in time (one subframe).[4] However, as will be described in Section 11.2.1, two of these symbols are used for *uplink demodulation reference signals* and are thus not available for PUSCH transmission. Furthermore, one additional symbol may be reserved for the transmission of *sounding reference signals*, see Section 11.2.2. Thus, 11 or 12 OFDM symbols are available for PUSCH transmission within each subframe.

Figure 11.3 illustrates how $K \cdot M$ DFT-precoded symbols at the output of the antenna mapping are mapped to the basic OFDM time–frequency grid, where K is the number of available OFDM symbols within a subframe (11 or 12 according to above) and M is the assigned bandwidth in number of subcarriers. As there are 12 subcarriers within a resource-block pair, $M = N \cdot 12$ where N is the number of assigned resource-block pairs.

Figure 11.3 assumes that the set of assigned resource-block pairs are contiguous in the frequency domain. This is the typical assumption for DFTS-OFDM and was strictly the case for LTE releases 8 and 9. Mapping of the DFT-precoded signal to frequency-contiguous resource elements is preferred in order to retain good cubic-metric properties of the uplink transmission. At the same time, such a restriction implies an additional constraint for the uplink scheduler, something which may not always be desirable. Therefore, LTE release 10 introduced the possibility to assign partly frequency-separated resources for PUSCH transmission. More specifically, in release 10 the assigned uplink resource may consist of a maximum of two frequency-separated *clusters* as illustrated in Figure 11.4, where each cluster consists of a number of resource-block pairs (N_1 and N_2 resource-block pairs respectively). In the case of such *multi-cluster transmission*, a single DFT precoding spans the overall assigned resource in the frequency domain—that is, both clusters. This means that the total assigned bandwidth in number of subcarriers ($M = M1 + M2$) should be aligned with the restrictions on available DFT sizes described above.

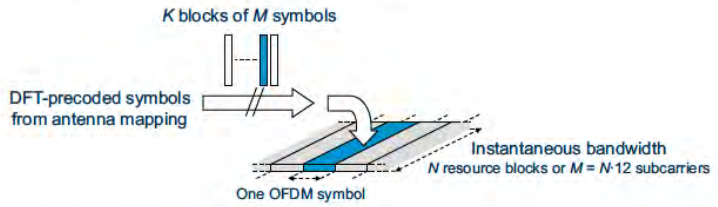


FIGURE 11.3

Mapping to the uplink physical resource

[4] 12 symbols in the case of extended cyclic prefix.

Copyright Elsevier 2024

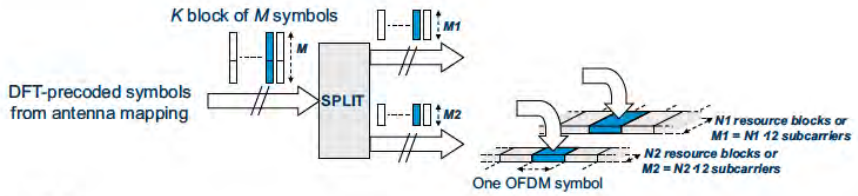


FIGURE 11.4

Uplink multi-cluster transmission

11.1.3 PUSCH frequency hopping

In Chapter 10 it was described how the notion of *virtual resource blocks* (VRBs) in combination with the mapping from VRBs to *physical resource blocks* (PRBs) allowed for downlink distributed transmission—that is, the spreading of a downlink transmission in the frequency domain. As described, downlink distributed transmission consists of two separate steps: (1) a mapping from VRB pairs to PRB pairs such that frequency-consecutive VRB pairs are not mapped to frequency-consecutive PRB pairs, and (2) a split of each resource-block pair such that the two resource blocks of the resource-block pair are transmitted with a certain frequency gap in between. This second step can be seen as frequency hopping on a slot basis.

The notion of VRBs can also be used for the LTE uplink, allowing for frequency-domain-distributed transmission for the uplink. However, in the uplink, where transmission from a terminal should always be over a set of consecutive subcarriers in the absence of multi-cluster transmission, distributing resource-block pairs in the frequency domain, as in the first step of downlink distributed transmission, is not possible. Rather, uplink distributed transmission is similar to the second step of downlink distributed transmission—that is, a frequency separation of the transmissions in the first and second slots of a subframe. Uplink distributed transmission for PUSCH can thus more directly be referred to as *uplink frequency hopping*.

There are two types of uplink frequency hopping defined for PUSCH:

- Sub-band-based hopping according to cell-specific hopping/mirroring patterns.
- Hopping based on explicit hopping information in the scheduling grant.

Uplink frequency hopping is not supported for multi-cluster transmission as, in that case, sufficient diversity is assumed to be obtained by proper location of the two clusters.

11.1.3.1 Hopping based on cell-specific hopping/mirroring patterns

To support sub-band-based hopping according to cell-specific hopping/mirroring patterns, a set of consecutive sub-bands of a certain size is defined from the overall uplink frequency band, as illustrated in Figure 11.5. It should be noted that the sub-bands do not cover the total uplink frequency band, mainly due to the fact that a number of resource blocks at the edges of

Copyright Elsevier 2024

Samsung Exhibit 1017, Page 271 of 538
Samsung Electronics Co., Ltd. v. Asus Technology Licensing Inc.
IPR2024-01217

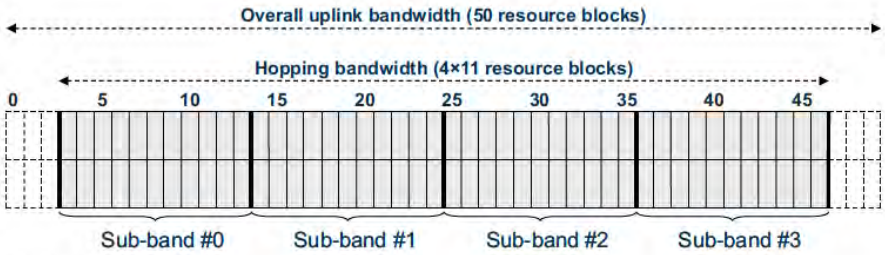


FIGURE 11.5

Definition of sub-bands for PUSCH hopping. The figure assumes a total of four sub-bands, each consisting of eleven resource blocks

the uplink frequency band are used for transmission of L1/L2 control signaling on the PUCCH. For example, in Figure 11.5, the overall uplink bandwidth corresponds to 50 resource blocks and there are a total of four sub-bands, each consisting of 11 resource blocks. Six resource blocks are not included in the hopping bandwidth and could, for example, be used for PUCCH transmission.

In the case of sub-band-based hopping, the set of virtual resource blocks provided in the scheduling grant are mapped to a corresponding set of physical resource blocks according to a cell-specific hopping pattern. The resource to use for transmission, the *PRBs*, is obtained by shifting the VRBs provided in the scheduling grant by a number of sub-bands according to the hopping pattern, where the hopping pattern can provide different shifts for each slot. As illustrated in Figure 11.6, a terminal is assigned VRBs 27, 28, and 29. In the first slot, the predefined hopping pattern takes the value 1, implying transmission using PRBs one sub-band to the right—that is, PRBs 38, 39, and 40. In the second slot, the predefined hopping pattern

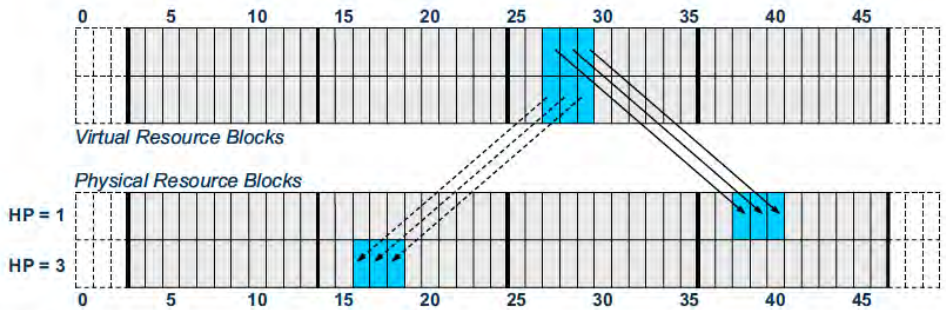


FIGURE 11.6

Hopping according to predefined hopping pattern

Copyright Elsevier 2024

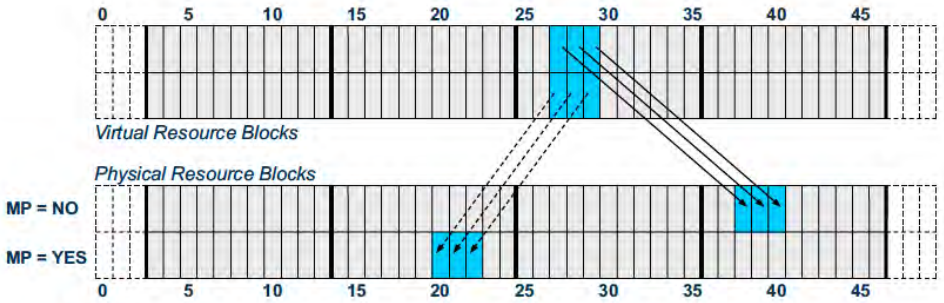


FIGURE 11.7

Hopping/mirroring according to predefined hopping/mirroring patterns. Same hopping pattern as in Figure 11.6

takes the value 3, implying a shift of three sub-bands to the right in the figure and, consequently, transmission using resource blocks 16, 17, and 18. Note that the shifting "wraps-around"—that is, in the case of four sub-bands, a shift of three sub-bands is the same as a negative shift of one sub-band. As the hopping pattern is cell specific the same for all terminals within a cell—different terminals will transmit on non-overlapping physical resources as long as they are assigned non-overlapping virtual resources.

In addition to the hopping pattern, there is also a cell-specific *mirroring pattern* defined in a cell. The mirroring pattern controls, on a slot basis, whether or not mirroring within each sub-band should be applied to the assigned resource. In essence, mirroring implies that the resource blocks within each sub-band are numbered right to left instead of left to right. Figure 11.7 illustrates mirroring in combination with hopping. Here, the mirroring pattern is such that mirroring is not applied to the first slot while mirroring is applied to the second slot.

Both the hopping pattern and the mirroring pattern depend on the physical-layer cell identity and are thus typically different in neighboring cells. Furthermore, the period of the hopping/mirroring patterns corresponds to one frame.

11.1.3.2 Hopping based on explicit hopping information

As an alternative to hopping/mirroring according to cell-specific hopping/mirroring patterns as described above, uplink slot-based frequency hopping for PUSCH can also be controlled by *explicit hopping information* provided in the scheduling grant. In such a case the scheduling grant includes:

- information about the resource to use for uplink transmission in the first slot, exactly as in the non-hopping case; and
- additional information about the offset of the resource to use for uplink transmission in the second slot, relative to the resource of the first slot.

Copyright Elsevier 2024

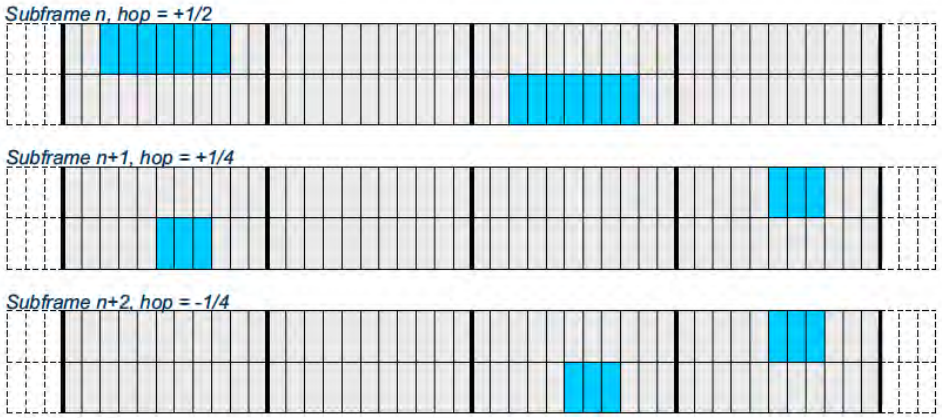


FIGURE 11.8

Frequency hopping according to explicit hopping information

Selection between hopping according to cell-specific hopping/mirroring patterns as described above or hopping according to explicit information in the scheduling grant can be done dynamically. More specifically, for cell bandwidths less than 50 resource blocks, there is a single bit in the scheduling grant indicating if hopping should be according to the cell-specific hopping/mirroring patterns or should be according to information in the scheduling grant. In the latter case, the hop is always half of the hopping bandwidth. In the case of larger bandwidths (50 resource blocks and beyond), there are two bits in the scheduling grant. One of the combinations indicates that hopping should be according to the cell-specific hopping/mirroring patterns while the three remaining alternatives indicate hopping of 1/2, +1/4, and −1/4 of the hopping bandwidth. Hopping according to information in the scheduling grant for the case of a cell bandwidth corresponding to 50 resource blocks is illustrated in Figure 11.8. In the first subframe, the scheduling grant indicates a hop of one-half the hopping bandwidth. In the second subframe, the grant indicates a hop of one-fourth the hopping bandwidth (equivalent to a negative hop of three-fourths of the hopping bandwidth). Finally, in the third subframe, the grant indicates a negative hop of one-fourth the hopping bandwidth.

11.2 Uplink reference signals

Similar to the downlink, reference signals are also transmitted on the LTE uplink. There are two types of reference signals defined for the LTE uplink:

- Uplink *demodulation reference signals* (DM-RS) are intended to be used by the base station for channel estimation for coherent demodulation of the uplink physical channels (PUSCH and PUCCH). A demodulation reference signal is thus only transmitted

Copyright Elsevier 2024

together with PUSCH or PUCCH and is then spanning the same frequency range as the corresponding physical channel.
- Uplink *sounding reference signals* (SRS) are intended to be used by the base station for channel-state estimation to support uplink channel-dependent scheduling and link adaptation. The SRS can also be used in other cases when uplink transmission is needed although there is no data to transmit. One example is when uplink transmission is needed for the network to be able to estimate the uplink receive timing as part of the *uplink-timing-alignment procedure* (see Section 11.6).

11.2.1 Demodulation reference signals

Uplink DM-RS are intended to be used for channel estimation for coherent demodulation of the Physical Uplink Shared Channel (PUSCH) to which the UL-SCH transport channel is mapped, as well as for the Physical Uplink Control Channel (PUCCH) which carries different types of uplink L1/L2 control signaling. The basic principles for uplink DM-RS are the same for PUSCH and PUCCH transmission although there are some differences—for example, in terms of the exact set of OFDM symbols in which the reference signals are transmitted. The discussion below will primarily focus on PUSCH DM-RM. Some additional details on the PUCCH DM-RS structure are provided in Section 11.4 as part of the more general description of uplink L1/L2 control signaling.

11.2.1.1 Time-frequency structure

Due to the importance of low cubic metric and corresponding high power-amplifier efficiency for uplink transmissions, the principles for uplink reference-signal transmission are different compared to the downlink. In essence, transmitting reference signals frequency multiplexed with other uplink transmissions from the same terminal is not suitable for the uplink as that would negatively impact the terminal power-amplifier efficiency due to increased cubic metric. Instead, certain OFDM symbols within a subframe are used exclusively for DM-RS transmission—that is, the reference signals are *time multiplexed* with other uplink transmissions (PUSCH and PUCCH) from the same terminal. The structure of the reference signal itself then ensures a low cubic metric within these symbols, as described below.

More specifically, in case of PUSCH transmission DM-RS is transmitted within the fourth symbol of each uplink slot[5] (Figure 11.9). Within each subframe, there are thus two reference-signal transmissions, one in each slot.

In case of PUCCH transmission, the number of OFDM symbols used for DM-RS transmission in a slot, as well as the exact position of these symbols, differ between different *PUCCH formats* as further described in Section 11.4.

In general, there is no reason to estimate the channel outside the frequency band of the corresponding PUSCH/PUCCH transmission that is to be coherently demodulated. The

[5]The third symbol in the case of extended cyclic prefix.

Copyright Elsevier 2024

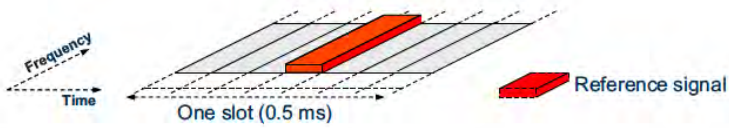


FIGURE 11.9

Transmission of uplink demodulation reference signals within a slot in case of PUSCH transmission

frequency range spanned by an uplink DM-RS is therefore equal to the instantaneous frequency range spanned by the corresponding PUSCH/PUCCH transmission. This means that, for PUSCH transmission, it should be possible to generate reference signals of different bandwidths, corresponding to the possible bandwidths of a PUSCH transmission. More specifically, it should be possible to generate reference signals of a bandwidth corresponding to $12 \cdot N$ subcarriers, where N corresponds to the bandwidth of the PUSCH transmission measured in number of resource blocks.[6]

As will be described in Section 11.4, PUCCH transmission from a terminal is always carried out over a single resource block in the frequency domain. In the case of PUCCH transmission, the bandwidth of the reference signal thus always corresponds to 12 subcarriers.

Regardless of the kind of uplink transmission (PUSCH or PUCCH), the basic structure of each reference-signal transmission is the same. As illustrated in Figure 11.10, an uplink DM-RS can be defined as a *frequency-domain reference-signal sequence* applied to consecutive inputs of an OFDM modulator—that is, to consecutive subcarriers. Referring to the previous discussion, in case of PUSCH transmission, the frequency-domain reference-signal sequence should have a length $M = 12 \cdot N$, where N corresponds to the PUSCH

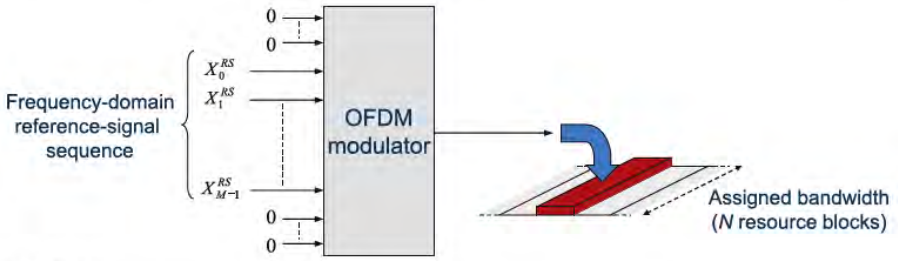


FIGURE 11.10

Generation of uplink DM-RS from a frequency-domain reference-signal sequence

[6]Due to the imposed limitations on supported DFT sizes as described in Section 11.1.1 there will be some additional constraints on N.

Copyright Elsevier 2024

bandwidth measured in number of resource blocks. In case of PUCCH transmission, the length of the reference-signal sequence should always be equal to 12.

11.2.1.2 Base sequences

Uplink reference signals should preferably have the following properties:

- Small power variations in the frequency domain to allow for similar channel-estimation quality for all frequencies spanned by the reference signal. Note that this is equivalent to a well-focused time-domain auto-correlation of the transmitted reference signal.
- Limited power variations in the time domain, leading to low cubic metric of the transmitted signal.

Furthermore, a sufficient number of reference-signal sequences of a given length, corresponding to a certain reference-signal bandwidth, should be available in order to avoid an unreasonable planning effort when assigning reference-signal sequences to terminals and cells.

So-called *Zadoff–Chu* sequences [71] have the property of constant power in both the frequency and time domains. The M_{ZC} elements of the q:th Zadoff–Chu sequence within the set of Zadoff–Chu sequences of length M_{ZC} can be expressed as:

$$Z_k^q = e^{\,j\pi q \frac{k\cdot(k+1)}{M_{ZC}}} \quad 0 \le k < M_{ZC} \tag{11.1}$$

From the point of view of small power variations in both the frequency and time domains, Zadoff–Chu sequences would thus be excellent as uplink reference-signal sequences. However, there are two reasons why Zadoff–Chu sequences are not suitable for direct use as uplink reference-signal sequences in LTE:

- The number of available Zadoff–Chu sequences of a certain length, corresponding to the number of possible values for the parameter q in expression (11.1) above, equals the number of integers that are relative prime to the sequence length M_{ZC}. To maximize the number of Zadoff–Chu sequences and thus, in the end, to maximize the number of available uplink reference signals, prime-length Zadoff–Chu sequences would therefore be preferred. At the same time, according to above, the length of the uplink reference-signal sequences should be a multiple of 12, which is not a prime number.
- For short sequence lengths, corresponding to narrow uplink transmission bandwidths, relatively few reference-signal sequences would be available even if they were based on prime-length Zadoff–Chu sequences.

Instead, for sequence lengths larger than or equal to 36, corresponding to transmission bandwidths larger than or equal to three resource blocks, basic reference-signal sequences, in the LTE specification referred to as *base sequences,* are defined as *cyclic extensions* of Zadoff–Chu sequences of length M_{ZC} (Figure 11.11), where M_{ZC} is the largest prime number smaller than or equal to the length of the reference-signal sequence. For example, the largest prime number less than or equal to 36 is 31, implying that reference-signal sequences of length 36 are defined as cyclic extensions of Zadoff–Chu sequences of length 31. The number

Copyright Elsevier 2024

Samsung Exhibit 1017, Page 277 of 538
Samsung Electronics Co., Ltd. v. Asus Technology Licensing Inc.
IPR2024-01217

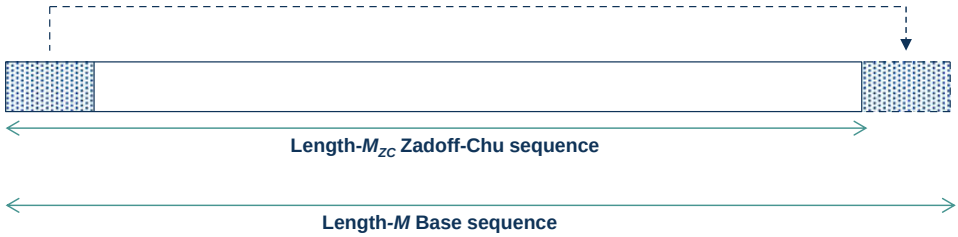


FIGURE 11.11

Length-M basic reference-signal sequence derived from cyclic extension of a length-M_{ZC} Zadoff-Chu sequence

of available sequences is then equal to 30—that is, one less than the length of the Zadoff–Chu sequence. For larger sequence lengths, more sequences are available. For example, for a sequence length equal to 72, there are 70 sequences available.[7]

For sequence lengths equal to 12 and 24, corresponding to transmission bandwidths of one and two resource blocks respectively, special QPSK-based sequences have instead been found from computer searches and are explicitly listed in the LTE specifications. For each of the two sequence lengths, 30 sequences have been defined.

Thus there are at least 30 sequences available for each sequence length. However, not all of these sequences are actually being used as base sequences.

- For sequence lengths less than 72, corresponding to reference-signal bandwidths less than six resource blocks, 30 sequences are being used.
- For sequence lengths equal to 72 and beyond, corresponding to reference-signal bandwidths of six resource blocks and beyond, 60 sequences are being used.

These sequences are divided into 30 *sequence groups* where each group consists of one base sequence for each sequence length less than 72 and two base sequences for each sequence length equal to 72 and beyond. A base sequence of a given length is thus fully specified by a *group index* ranging from 0 to 29 together with, in case of sequence lengths equal to 72 and beyond, a *sequence index* taking the values 0 and 1.

11.2.1.3 Phase-rotation and orthogonal cover codes

From the base sequences previously described, additional reference-signal sequences can be generated by applying different linear phase rotations in the frequency domain, as illustrated in Figure 11.12.

Applying a linear phase rotation in the frequency domain is equivalent to applying a cyclic shift in the time domain. Thus, although being *defined* as different frequency-domain phase rotations in line with Figure 11.12, the LTE specification actually refers to this as applying different *cyclic*

[7]The largest prime number smaller than or equal to 72 is 71. The number of sequences is then one less than the length of the Zadoff–Chu sequence, i.e., 70.

Copyright Elsevier 2024

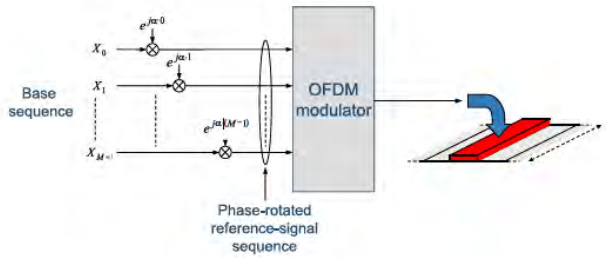


FIGURE 11.12

Generation of uplink reference-signal sequence from phase-rotated base sequence

shifts. Here the term "phase rotation" will be used. However, it should be borne in mind that what is here referred to as phase rotation is referred to as cyclic shift in the 3GPP specifications.

Demodulation reference signals derived from different base sequences typically have relatively low but still non-zero cross correlation. In contrast, reference signals defined from different phase rotations of the same base sequence are, at least in theory, completely orthogonal if the parameter α in Figure 11.12 takes a value $m\pi/6$, where the integer m ranges from 0 to 11.[8] Up to 12 orthogonal reference signals can thus be derived from each base sequence by using different values of the parameter m.

However, to retain the orthogonality between these reference signals at the receiver side, the frequency response of the channel should be essentially constant over the span of one resource block—that is, over 12 subcarriers. Alternatively expressed, the main part of the channel time dispersion should not extend beyond the length of the cyclic shift mentioned above. If that is not the case, a subset of the available values for α may be used—for example, only the values $\{0, 2\pi/6, 4\pi/6, \ldots, 10\pi/6\}$ or perhaps even fewer values. Limiting the set of possible values for α implies orthogonality over a smaller number of subcarriers and, as a consequence, less sensitivity to channel frequency selectivity. In other words, there is a trade-off between the number of orthogonal reference signals that can be generated from different phase rotations and the amount of channel frequency-selectivity that should be possible to cope with.

Another prerequisite for receiver-side orthogonality between reference signals defined from different phase rotations of the same base sequence is that the reference signals are transmitted well aligned in time. Thus the main uses of phase rotation include:

- To provide multiple simultaneous reference signals from the same terminal in case of uplink multi-layer transmission (uplink spatial multiplexing, see Section 11.3.1)
- To provide the possibility for orthogonal reference signals between multiple terminals being scheduled for PUSCH transmission on the same resource, that is, the same set of resource blocks, within a cell (uplink MU-MIMO, see Section 11.3.2)

[8]The orthogonality is due to the fact that, for $\alpha = m\pi/6$, there will be an integer number of full-circle rotations over 12 subarriers—that is, over one resource block.

Copyright Elsevier 2024

Phase rotations may also be used to provide orthogonal reference signals between terminals in neighbor cells, assuming tight synchronization between the cells. Finally, phase rotations are also used to separate reference signals of different terminals in case of PUCCH transmission (see Section 11.4).

In addition to the use of different phase rotations, orthogonal reference-signal transmissions can also be achieved by means of *Orthogonal Cover Codes* (OCC). As illustrated in Figure 11.13, two different length-two OCCs ([+1 +1] and [+1 −1] respectively) can be applied to the two PUSCH reference-signal transmissions within a subframe. This allows for overall DM-RS orthogonality over the subframe assuming that

- the channel does not change substantially over the subframe and
- the reference signals of the two slots are the same.[9]

Similar to phase rotations, receiver-side orthogonality between reference-signal transmissions based on different OCC requires that the transmissions are well-aligned in time at the receiver side. Thus the use of OCC is essentially the same as for phase rotations as described earlier:

- To provide multiple reference signals from the same terminal in case of uplink spatial multiplexing
- To provide orthogonal reference signals between terminals being scheduled on the same resource within a cell (uplink MU-MIMO)
- To allow for reference-signal orthogonality between uplink transmissions within neighbor cells in case of tight synchronization and time alignment between the cells

It should be noted that, in contrast to phase rotations, orthogonality by means of OCC does not require that the same base sequence is used for the two DM-RS (DM-RS 0 and

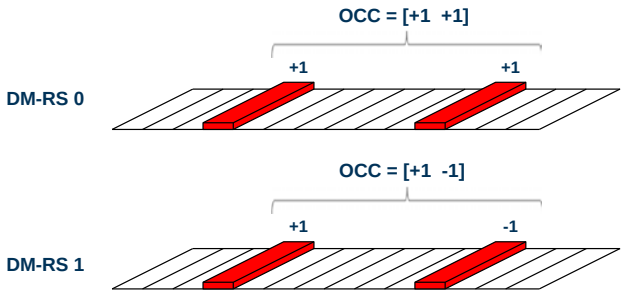


FIGURE 11.13

Generation of multiple demodulation reference signals from orthogonal cover codes

[9]Strictly speaking, the only thing that is required is that the correlation between the reference signals of DM-RS 0 and DM-RS 1 in Figure 11.14 is the same for the two slots. If the reference signals are the same for the two slots, this is the case.

Copyright Elsevier 2024

DM-RS 1 in Figure 11.13). Actually, the two reference signals do not even need to have the same bandwidth; having the same cross-correlation between the reference signals of DM-RS 0 and DM-RS 1 for the two slots is sufficient. Thus OCC can be used to achieve reference-signal orthogonality also for PUSCH transmissions of different bandwidths.

Similar to phase rotations, orthogonal codes can also be applied to DM-RS in case of PUCCH transmission, although in a somewhat different way compared to PUSCH due to the different time-domain structure of PUCCH DM-RS (see Section 11.4).

11.2.1.4 Base-sequence assignment

According to above each base sequence of a given length corresponds to a unique combination of a group index ranging from 0 to 29 and a sequence index taking the values 0 or 1. Base-sequence assignment—that is, determining what base sequence should be used by a specific terminal—is thus equivalent to assigning a corresponding group and sequence index.[10]

Prior to release 11, base-sequence assignment was *cell-specific*—that is, for a given slot the group and sequence indices were the same for all terminals having the same serving cell.

In the case of a *fixed* (*non-hopping*) group assignment, the sequence group to use for PUCCH transmission does not change between slots and was, prior to release 11, directly given by the physical-layer cell identity. More specifically, the group index was equal to the cell identity modulo 30, where the cell identity may take values in the range 0 to 503 as described in Chapter 14. Thus, cell identities 0, 30, 60, …, 480 corresponded to sequence group 0, cell identities 1, 31, 61, …, 481 to sequence group 1, and so on.

In contrast, what sequence group to use for PUSCH transmission could be explicitly configured on a cell basis by adding an offset provided as part of the cell system information to the PUCCH group index. The reason for providing the possibility for explicitly indicating what sequence group to use for PUSCH transmission in a cell was that it should be possible to use the same sequence group for PUSCH transmission in neighboring cells, despite the fact that such cells typically have different cell identities. In this case, the reference signals for PUSCH transmissions within the two cells would instead be distinguished by different phase rotations and/or OCC as discussed in Section 11.2.1.3, allowing for reference-signal orthogonality also *between* cells.[11]

In the case of *group hopping*, an additional cell-specific *group-hopping pattern* is added to the group index, allowing for the group index of a cell to change on a slot basis. Prior to release 11, the group-hopping pattern was also derived from the cell identity, and identical group-hopping patterns were used for PUSCH and PUCCH within a cell.

In addition to the group index, for sequence lengths equal to or larger than 72, the reference-signal sequence also depends on the *sequence index*. The sequence index can either be fixed (non-hopping), in which case it always equals 0, or vary between slots

[10]For base-sequence lengths less than 72, the sequence index is always equal to zero.
[11]This assumes tight synchronization and time-alignment between the cells.

Copyright Elsevier 2024

(hopping) according to a *sequence-hopping pattern*. Similar to group hopping, prior to release 11 the sequence-hopping pattern was cell-specific and given by the physical-layer cell identity.

In LTE release 11, the possibility for *terminal-specific* base sequence assignment was introduced—that is, the group and sequence indices to use for PUSCH and PUCCH can, with release 11, be explicitly configured for a specific terminal regardless of the identity of the serving cell. The introduction of terminal-specific base-sequence assignment was done very similarly to the introduction of terminal-specific downlink DM-RS as described in Section 10.2.2—that is, by introducing the possibility for explicitly configuring a terminal with something that can be seen as a "virtual cell identity" which, if configured, replaces that actual physical-cell identity when deriving the group and sequence index. Similar to downlink DM-RS, if no virtual cell identity is configured the terminal should assume cell-specific base-sequence assignment according to above.

It should be pointed out that the terminal is not configured with a single "virtual-cell identity" to be used for deriving the base sequences for PUSCH and PUCCH. Rather, the terminal-specific configuration is done separately for PUSCH and PUCCH using two different "virtual cell identities."

The main reason for introducing the possibility for full terminal-specific assignment of uplink reference signals was to enhance the support for uplink *multi-point reception*. In case of uplink multi-point reception, an uplink transmission may be received at a reception point not corresponding to the serving cell of the terminal but at a reception point corresponding to another cell. To enhance the reception quality it is beneficial to have orthogonal reference signals for different terminals received at this reception point despite the fact that they, strictly speaking, have different serving cells. To allow for this, the reference-signal sequences should not be cell-specific but rather possible to assign on a terminal basis. Multi-point reception is discussed in somewhat more detail in Chapter 15 as part of a more general discussion on multi-point coordination and transmission.

11.2.1.5 Assignment of phase rotation and OCC

As discussed previously, the main use for phase rotations and OCC in case of PUSCH DM-RS is to provide a possibility for orthogonal reference signals for different layers in case of spatial multiplexing and for different terminals scheduled on the same resource either within one cell (MU-MIMO) or in neighboring tightly synchronized cells.

In order not to limit the flexibility in terms of what terminals can be jointly scheduled on the same resource, the assignment of phase rotation and OCC can be done dynamically. Thus, the exact phase rotation, given by the phase parameter m in Figure 11.12 and the OCC is provided jointly as a single parameter included as part of the uplink scheduling grant provided by the network. Each value of this parameter corresponds to a certain combination of phase rotation and OCC for each layer to be transmitted by the terminal. In case of spatial multiplexing, the different layers will then inherently be assigned different phase shifts and, possibly, different OCC. By providing different parameter values to different terminals, the

Copyright Elsevier 2024

terminals will be assigned different phase-shifts/OCC combinations, allowing for orthogonal reference signals and thus providing enhanced MU-MIMO performance either within a cell or between cells.

11.2.2 Sounding reference signals

The demodulation reference signals discussed in Section 11.2.1 are intended to be used by the base station for channel estimation to allow for coherent demodulation of uplink physical channels (PUSCH or PUCCH). A demodulation reference signal is always transmitted together with and spanning the same frequency range as the corresponding physical channel.

In contrast, *sounding reference signals* (SRS) are transmitted on the uplink to allow for the base station to estimate the uplink *channel state* at different frequencies. The channel-state estimates can then, for example, be used by the base-station scheduler to assign resource blocks of instantaneously good quality for uplink PUSCH transmission from the specific terminal (uplink channel-dependent scheduling), as well as to select different transmission parameters such as the instantaneous data rate and different parameters related to uplink multi-antenna transmission. As mentioned earlier, SRS transmission can also be used in other situations when uplink transmission is needed although there is no data to transmit, for example, for uplink timing estimation as part of the uplink-timing-alignment procedure. Thus, an SRS is not necessarily transmitted together with any physical channel and if transmitted together with, for example, PUSCH, the SRS may span a different, typically larger, frequency range.

There are two types of SRS transmission defined for the LTE uplink: *periodic* SRS transmission, which has been available from the first release of LTE (release 8) and *aperiodic* SRS transmission, introduced in LTE release 10.

11.2.2.1 Periodic SRS transmission

Periodic SRS transmission from a terminal occurs at regular time intervals, from as often as once every 2 ms (every second subframe) to as infrequently as once every 160 ms (every 16th frame). When SRS is transmitted in a subframe, it occupies the last symbol of the subframe as illustrated in Figure 11.14. As an alternative, in the case of TDD operation, SRS can also be transmitted within the UpPTS.

In the frequency domain, SRS transmissions should span the frequency range of interest for the scheduler. This can be achieved in two ways:

- By means of a sufficiently wideband SRS transmission that allows for sounding of the entire frequency range of interest with a single SRS transmission, as illustrated in the upper part of Figure 11.15.
- By means of a more narrowband SRS transmission hopping in the frequency domain in such a way that a sequence of SRS transmissions jointly spans the frequency range of interest, as illustrated in the lower part of Figure 11.15.

Copyright Elsevier 2024

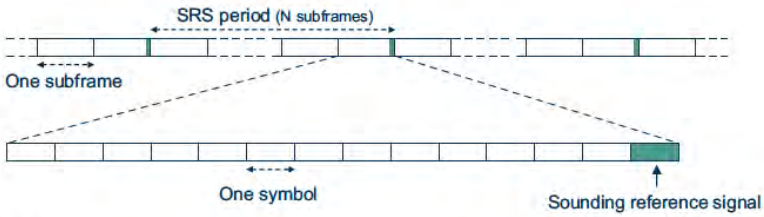


FIGURE 11.14

Transmission of SRS

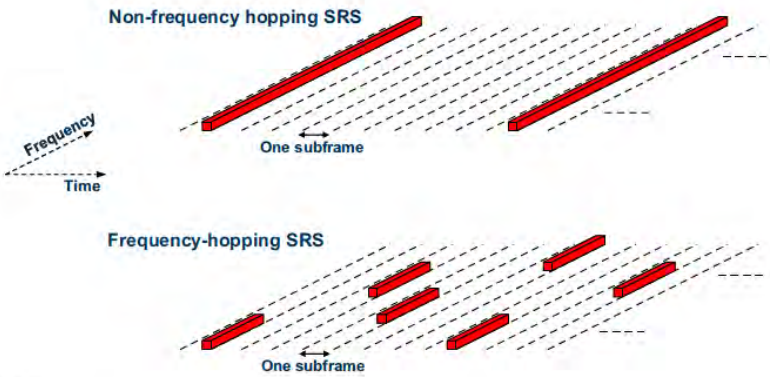


FIGURE 11.15

Non-frequency-hopping (wideband) versus frequency-hopping SRS

The instantaneous SRS bandwidth is always a multiple of four resource blocks. Different bandwidths of the SRS transmission can simultaneously be available within a cell. A narrow SRS bandwidth, corresponding to four resource blocks, is always available, regardless of the uplink cell bandwidth. Up to three additional, more wideband SRS bandwidths may also be configured within a cell. A specific terminal within the cell is then explicitly configured to use one of up to four SRS bandwidths available in the cell.

If a terminal is transmitting SRS in a certain subframe, the SRS transmission may very well overlap, in the frequency domain, with PUSCH transmissions from other terminals within the cell. To avoid collision between SRS and PUSCH transmissions from different terminals, terminals should in general avoid PUSCH transmission in the OFDM symbols in which SRS transmission may occur. To achieve this, all terminals within a cell are aware of the set of subframes within which SRS *may* be transmitted by *any* terminal within the cell.

Copyright Elsevier 2024

All terminals should then avoid PUSCH transmission in the last OFDM symbol of those subframes. Information about the set of subframes in which SRS *may be* transmitted within a cell is provided as part of the cell system information.[12]

On a more detailed level, the structure for sounding reference signals (SRS) is similar to that of uplink demodulation reference signals described in Section 11.2.1. More specifically, a sounding reference signal is also defined as a frequency-domain reference-signal sequence derived in the same way as for DM-RS—that is, cyclic extension of prime-length Zadoff–Chu sequences for sequence lengths equal to 30 and above and special sequences for sequence lengths less than 30. However, in the case of SRS, the reference-signal sequence is mapped to *every second* subcarrier, creating a "comb"-like spectrum, as illustrated in Figure 11.16. Taking into account that the bandwidth of the SRS transmission is always a multiple of four resource blocks, the lengths of the reference-signal sequences for SRS are thus always a multiple of 24.[13] The reference-signal sequence to use for SRS transmission within the cell is derived in the same way as the PUCCH DM-RS within the cell, assuming cell-specific reference-signal sequence assignment. Terminal-specific reference-signal sequences are not supported for sounding reference signals.

Similar to demodulation reference signals, different phase rotations, also for SRS referred to as "cyclic shifts" in the LTE specifications, can be used to generate different SRS that are orthogonal to each other. By assigning different phase rotations to different terminals, multiple SRS can thus be transmitted in parallel in the same subframe, as illustrated by terminals #1 and #2 in the upper part of Figure 11.17. However, it is then required that the reference signals span the same frequency range.

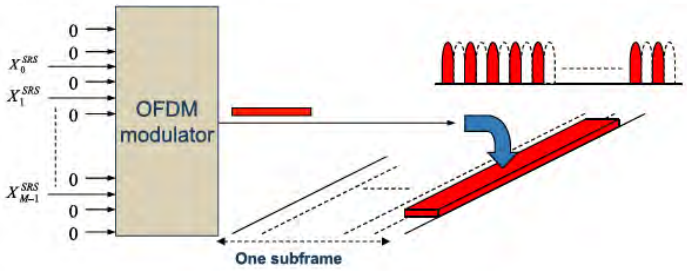


FIGURE 11.16

Generation of SRS from a frequency-domain reference-signal sequence

[12]What is provided as part of the system information is a periodicity (2 to 160 ms) and a subframe offset; compare the bullet list in the following.

[13]Four resource blocks, each spanning 12 subcarriers, but only every second subcarrier is used for a certain SRS transmission.

Copyright Elsevier 2024

Samsung Exhibit 1017, Page 285 of 538
Samsung Electronics Co., Ltd. v. Asus Technology Licensing Inc.
IPR2024-01217

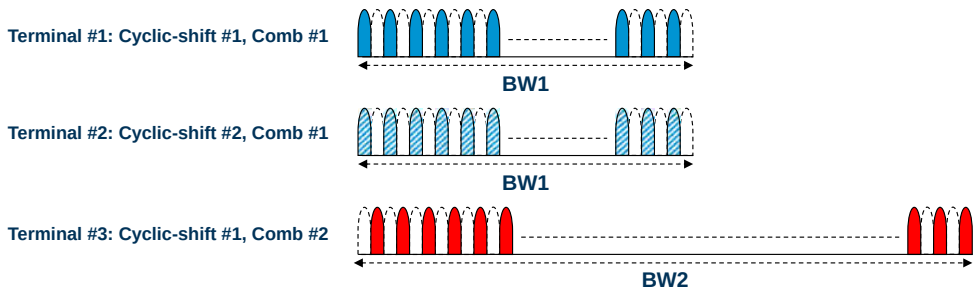


FIGURE 11.17

Multiplexing of SRS transmissions from different terminals

Another way to allow for SRS to be simultaneously transmitted from different terminals is to rely on the fact that each SRS only occupies every second subcarrier. Thus, SRS transmissions from two terminals can be *frequency multiplexed* by assigning them to different frequency shifts or "combs," as illustrated by terminal #3 in the lower part of Figure 11.17. In contrast to the multiplexing of SRS transmission by means of different "cyclic shifts," frequency multiplexing of SRS transmissions does not require the transmissions to cover identical frequency ranges.

To summarize, the following set of parameters defines the characteristics of an SRS transmission:

- SRS transmission time-domain period (from 2 to 160 ms) and subframe offset.
- SRS transmission bandwidth—the bandwidth covered by a single SRS transmission.
- Hopping bandwidth—the frequency band over which the SRS transmission is frequency hopping.
- Frequency-domain position—the starting point of the SRS transmission in the frequency domain.
- Transmission comb as illustrated in Figure 11.17.
- Phase rotation (or equivalently cyclic shift) of the reference-signal sequence.

A terminal that is to transmit SRS is configured with these parameters by means of higher layer (RRC) signaling.

11.2.2.2 Aperiodic SRS transmission

In contrast to periodic SRS, aperiodic SRS are *one-shot* transmissions, triggered by signaling on PDCCH as part of the scheduling grant. The frequency-domain structure of an aperiodic SRS transmission is identical to that of periodic SRS. Also, in the same way as for periodic SRS transmission, aperiodic SRS are transmitted within the last symbol of a subframe.

Copyright Elsevier 2024

Furthermore, the time instants when aperiodic SRS may be transmitted are configured per terminal using higher-layer signaling.

The frequency-domain parameters for aperiodic SRS (bandwidth, odd or even "comb," etc.) are configured by higher-layer (RRC) signaling. However, no SRS transmission will actually be carried out until the terminal is explicitly triggered to do so by an explicit *SRS trigger* on PDCCH/EPDCCH. When such a trigger is received, a single SRS is transmitted in the next available aperiodic SRS instant configured for the terminal using the configured frequency-domain parameters. Additional SRS transmissions can then be carried out if additional triggers are received.

Three different parameter sets can be configured for aperiodic SRS, for example differing in the frequency position of the SRS transmission and/or the transmission comb. Information on what parameters to use when the SRS is actually transmitted is included in the PDCCH/EPDCCH L1/L2 control-signaling information, which consists of two bits, three combinations of which indicate the specific SRS parameter set. The fourth combination simply indicates that no SRS should be transmitted.

11.3 Uplink multi-antenna transmission

Downlink multi-antenna transmission was supported by the LTE specification from its first release (release 8). With LTE release 10, support for uplink multi-antenna transmission—that is, uplink transmission relying on multiple transmit antennas at the terminal side—was also introduced for LTE. Uplink multi-antenna transmission can be used to improve the uplink performance in different ways:

- To improve the achievable data rates and spectral efficiency for uplink data transmission by allowing for antenna precoding supporting uplink beam-forming as well as spatial multiplying with up to four layers for the uplink physical data channel PUSCH.
- To improve the uplink control-channel performance by allowing for transmit diversity for the uplink physical control channel PUCCH.

11.3.1 Precoder-based multi-antenna transmission for PUSCH

As illustrated in Figure 11.18, the structure of the uplink antenna precoding is very similar to that of downlink antenna precoding (Section 10.3), including the presence of precoded demodulation reference signals (one per layer) similar to downlink non-codebook-based precoding (Figure 10.19). Uplink antenna precoding supports transmission using up to four antenna ports, allowing for spatial multiplexing with up to four layers.

The principles for mapping of the modulation symbols to layers are also the same as for the downlink. For an initial transmission, there is one transport block in the case of a single layer and two transport blocks for more than one layer, as illustrated in Figure 11.19. Similar

Copyright Elsevier 2024

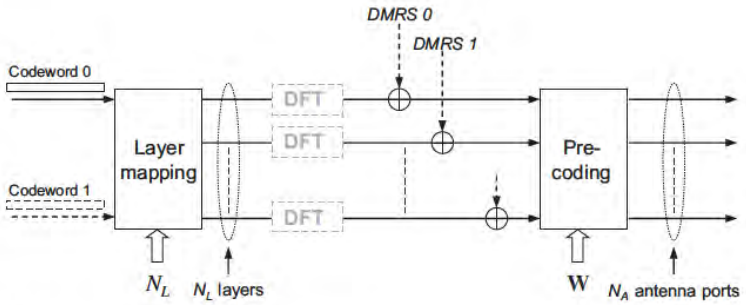


FIGURE 11.18

Precoder-based multi-antenna transmission for LTE uplink

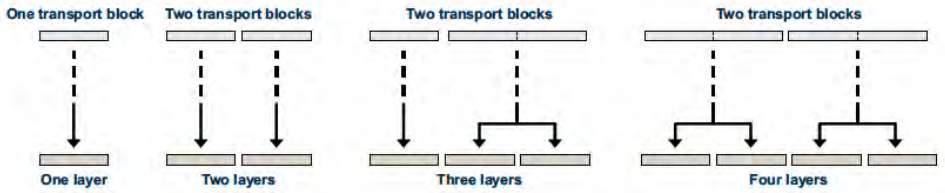


FIGURE 11.19

Uplink transport-channel-to-layer mapping (initial transmission)

to the downlink, in the case of a hybrid-ARQ retransmission, a single transport block may also be transmitted on multiple layers in some cases.

As can be seen in Figure 11.18, the DFT precoding is actually taking place after layer mapping—that is, each layer is separately DFT precoded. To simplify the description this was not really visible in Figure 11.1 outlining the overall physical-layer transport-channel processing.

It can also be noted that, in contrast to Figure 10.19, the precoder in Figure 11.18 is not shaded. As discussed in Section 10.3.3, for downlink non-codebook-based precoding, the precoder part of the antenna mapping is not visible in the specification and the network can, in essence, apply an arbitrary precoding for the downlink transmission. Due to the use of precoded demodulation reference signals, the terminal can recover the different layers without knowledge of exactly what precoding has been applied at the transmitted side.

The same is also true for the uplink—that is, the presence of precoded demodulation reference signals would allow for the base station to demodulate the uplink multi-antenna transmission and recover the different layers without knowledge of the precoding taking

Copyright Elsevier 2024

place at the transmitter side. However, for LTE the uplink precoder matrix is selected by the network and conveyed to the terminal as part of the scheduling grant. The terminal should then follow the precoder matrix selected by the network. Thus, in the uplink, the precoder is visible in the specification and, in order to limit the downlink signaling, there is a limited set of precoder matrices specified for each transmission rank.

More specifically, for each combination of transmission rank N_L and number of antennas ports N_A, a set of precoder matrices of size $N_A \times N_L$ is defined, as illustrated in Tables 11.1 and 11.2 for two and four antenna ports respectively. For full-rank transmission—that is, when the transmission rank or number of layers equals the number of transmit antennas—only a single precoder matrix is defined, namely the identity matrix of size $N_A \times N_A$ (not shown in the tables). Note that, for the case of four antenna ports, only a subset of the defined matrices is shown. In total there are 24 rank-1 matrices, 16 rank-2 matrices, and 12 rank-3 matrices defined for four antenna ports, in addition to the single rank-4 matrix.

As can be seen, all the precoder matrices in Table 11.1 contain one and only one non-zero element in each row, and this is generally true for all precoder matrices defined for the uplink. As a consequence, the signal transmitted on a certain antenna port (corresponding to a certain row of the precoder matrix) always depends on one and only one specific layer (corresponding to a specific column of the precoder matrix). Expressed alternatively, the precoder matrix maps the layers to the antenna ports *with at most one layer being mapped to each antenna port*. Due to this, the good cubic-metric properties of the transmitted signal are also preserved for each antenna port when antenna precoding is applied. The precoder matrices of Tables 11.1 and 11.2 are therefore also referred to as *cubic-metric-preserving precoder matrices*.

In order to select a suitable precoder, the network needs information about the uplink channel. Such information can, for example, be based on measurements on the uplink sounding reference signals (Section 11.2.2). As indicated in Figure 11.20, sounding reference signals (SRS) are transmitted non-precoded—that is, directly on the different antenna ports. The received SRS thus reflect the channel of each antenna port, not including any precoding.

Table 11.1 Uplink Precoder Matrices for two Antenna Ports

	Codebook Index					
Transmission Rank	**0**	**1**	**2**	**3**	**4**	**5**
1	$\frac{1}{\sqrt{2}}\begin{bmatrix}1\\1\end{bmatrix}$	$\frac{1}{\sqrt{2}}\begin{bmatrix}1\\-1\end{bmatrix}$	$\frac{1}{\sqrt{2}}\begin{bmatrix}1\\j\end{bmatrix}$	$\frac{1}{\sqrt{2}}\begin{bmatrix}1\\-j\end{bmatrix}$	$\frac{1}{\sqrt{2}}\begin{bmatrix}1\\0\end{bmatrix}$	$\frac{1}{\sqrt{2}}\begin{bmatrix}0\\1\end{bmatrix}$
2	$\frac{1}{\sqrt{2}}\begin{bmatrix}1&0\\0&1\end{bmatrix}$	-	-	-	-	-

Copyright Elsevier 2024

Table 11.2 Subset of Uplink Precoder Matrices for Four Antenna Ports and Different Transmission Ranks

Transmission Rank	Codebook Index 0	1	2	3	...
1	$\frac{1}{2}\begin{bmatrix}1\\1\\1\\-1\end{bmatrix}$	$\frac{1}{2}\begin{bmatrix}1\\1\\j\\j\end{bmatrix}$	$\frac{1}{2}\begin{bmatrix}1\\1\\-1\\1\end{bmatrix}$	$\frac{1}{2}\begin{bmatrix}1\\1\\-j\\-j\end{bmatrix}$	...
2	$\frac{1}{2}\begin{bmatrix}1&0\\1&0\\0&1\\0&-j\end{bmatrix}$	$\frac{1}{2}\begin{bmatrix}1&0\\1&0\\0&1\\0&j\end{bmatrix}$	$\frac{1}{2}\begin{bmatrix}1&0\\-j&0\\0&1\\0&1\end{bmatrix}$	$\frac{1}{2}\begin{bmatrix}1&0\\-j&0\\0&1\\0&-1\end{bmatrix}$	...
3	$\frac{1}{2}\begin{bmatrix}1&0&0\\1&0&0\\0&1&0\\0&0&1\end{bmatrix}$	$\frac{1}{2}\begin{bmatrix}1&0&0\\-1&0&0\\0&1&0\\0&0&1\end{bmatrix}$	$\frac{1}{2}\begin{bmatrix}1&0&0\\0&1&0\\1&0&0\\0&0&1\end{bmatrix}$	$\frac{1}{2}\begin{bmatrix}1&0&0\\0&1&0\\-1&0&0\\0&0&1\end{bmatrix}$	...
4	$\frac{1}{2}\begin{bmatrix}1&0&0&0\\0&1&0&0\\0&0&1&0\\0&0&0&1\end{bmatrix}$	-	-	-	...

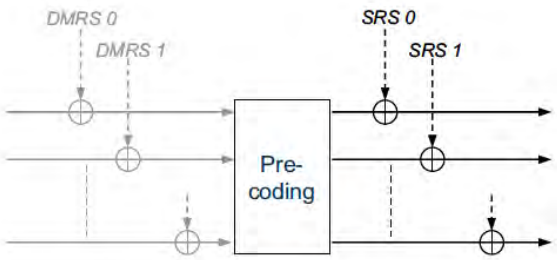


FIGURE 11.20

Illustration of SRS transmitted *after* uplink antenna precoding

Based on the received SRS, the network can thus decide on a suitable uplink transmission rank and corresponding uplink precoder matrix, and provide information about the selected rank and precoder matrix as part of the scheduling grant.

The previous paragraph assumed the same number of antenna ports for PUSCH as for SRS. This is a relevant situation and the SRS is, in this case, used to aid the selection of the

Copyright Elsevier 2024

precoding matrix, as discussed above. However, there are also situations when SRS and PUSCH use *different* numbers of antenna ports. One example is uplink transmission of two layers (two antenna ports), where the eNodeB would like to use SRS to probe the channel for potential four-layer transmission. In this case the SRS is transmitted on a *different* set of antenna ports than the PUSCH to aid the eNodeB in assessing the benefits, if any, of switching to four-layer transmission.

11.3.2 Uplink multi-user MIMO

As described in Section 10.3.5, downlink multi-user MIMO (MU-MIMO) implies downlink transmission to different terminals using *the same time–frequency resource* and relying on the availability of multiple antennas, at least on the network side, to suppress interference between the transmissions. The term MU-MIMO originated from the resemblance to SU-MIMO (spatial multiplexing).

Uplink MU-MIMO is essentially the same thing but for the uplink transmission direction—that is, uplink MU-MIMO implies uplink transmissions from multiple terminals using *the same uplink time–frequency resource* and relying on the availability of multiple receive antennas at the base station to separate the two or more transmissions. Thus, MU-MIMO is really just another term for uplink *Space-Division Multiple Access* (SDMA).

Actually, on the uplink, the relation between MU-MIMO and SU-MIMO (spatial multiplexing) is even closer. Uplink spatial multiplexing, for example with two antenna ports and two layers, implies that the terminal transmits two transport blocks with one transport block transmitted on each layer and thus on each antenna port,[14] as illustrated in the left part of Figure 11.21. As illustrated in the right part of the figure, MU-MIMO is essentially equivalent to separating the two antennas into two different terminals and transmitting one

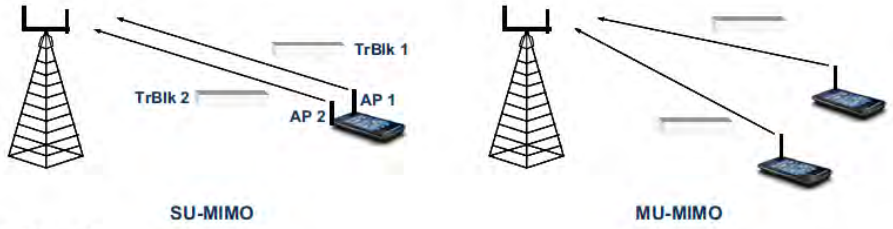


FIGURE 11.21

SU-MIMO and MU-MIMO

[14]Note that the 2×2 precoder matrix is the identity matrix (see Table 11.1).

Copyright Elsevier 2024

transport block from each terminal. The base-station processing to separate the two transmissions could essentially be identical to the processing used to separate the two layers in the case of spatial multiplexing. It should be noted that the separation of the two transmissions at the receiver side could be simplified, or at least the possible means to achieve this separation are extended, if the two terminals are well separated in space, something which is not the case for two antennas attached to the same terminal. As an example, for sufficiently separated terminals, classical beam-forming relying on correlated receiver antennas can be used to separate the uplink transmissions. Alternatively, uncorrelated receiver antennas can be used, and the separation means are then essentially the same as for SU-MIMO.

One important benefit of uplink MU-MIMO is that one, in many cases, can get similar gains in *system throughput* as SU-MIMO (spatial multiplexing) without the need for multiple transmit antennas at the terminal side, allowing for less complex terminal implementation. It should be noted though that spatial multiplexing could still provide substantial gains in terms of *user throughput* and peak data rates that can be provided from a single terminal. Furthermore, the potential system gains of uplink MU-MIMO rely on the fact that more than one terminal is actually available for transmission to in a subframe. The process of "pairing" terminals that should share the time–frequency resources is also non-trivial and requires suitable radio-channel conditions.

Essentially, support for uplink MU-MIMO only requires the possibility to explicitly assign a specific orthogonal reference signal for the uplink transmission, thereby ensuring orthogonality between reference-signal transmissions from the different terminals involved in the MU-MIMO transmission. As described in Section 11.2.1.5, this is supported by means of the dynamic assignment of DM-RS phase rotation and OCC as part of the uplink scheduling grant.

11.3.3 PUCCH transmit diversity

Precoder-based multi-layer transmission is only used for the uplink data transmission on PUSCH. However, in the case of a terminal with multiple transmit antennas, one wants to use the full set of terminal antennas and corresponding terminal power amplifiers also for the L1/L2 control signaling on PUCCH in order to be able to utilize the full power resource and achieve maximum diversity. To achieve additional diversity, LTE release 10 also introduced the possibility for two-antenna *transmit diversity* for PUCCH. More specifically, the transmit diversity supported for PUCCH is referred to as Spatial Orthogonal-Resource Transmit Diversity (SORTD).

The basic principle of SORTD is simply to transmit the uplink control signaling using different resources (time, frequency, and/or code) on the different antennas. In essence, the PUCCH transmissions from the two antennas will be identical to PUCCH transmissions from two different terminals using different resources. Thus, SORTD creates additional diversity but achieves this by using twice as many PUCCH resources, compared to non-SORTD transmission.

Copyright Elsevier 2024

For four physical antennas at the terminal, implementation-specific *antenna virtualization* is used. In essence, a transparent scheme is used to map the two-antenna-port signal to four physical antennas.

11.4 Uplink L1/L2 control signaling

Similar to the LTE downlink, there is also a need for uplink L1/L2 control signaling to support data transmission on downlink and uplink transport channels. Uplink L1/L2 control signaling consists of:

- hybrid-ARQ acknowledgements for received DL-SCH transport blocks;
- channel-state information (CSI) related to the downlink channel conditions, used to assist downlink scheduling; and
- scheduling requests, indicating that a terminal needs uplink resources for UL-SCH transmission.

There is no information indicating the UL-SCH transport format signaled on the uplink. As mentioned in Chapter 8, the eNodeB is in complete control of the uplink UL-SCH transmissions and the terminal always follows the scheduling grants received from the network, including the UL-SCH transport format specified in those grants. Thus, the network knows the transport format used for the UL-SCH transmission in advance and there is no need for any explicit transport-format signaling on the uplink.

Uplink L1/L2 control signaling needs to be transmitted on the uplink regardless of whether or not the terminal has any uplink transport-channel data to transmit and thus regardless of whether or not the terminal has been assigned any uplink resources for UL-SCH transmission. Hence, two different methods are supported for the transmission of the uplink L1/L2 control signaling, depending on whether or not the terminal has been assigned an uplink resource for UL-SCH transmission:

- *Non-simultaneous transmission of UL-SCH and L1/L2 control.* If the terminal does not have a valid scheduling grant—that is, no resources have been assigned for the UL-SCH in the current subframe—a separate physical channel, the *Physical Uplink Control Channel* (PUCCH), is used for transmission of uplink L1/L2 control signaling.
- *Simultaneous transmission of UL-SCH and L1/L2 control.* If the terminal has a valid scheduling grant—that is, resources have been assigned for the UL-SCH in the current subframe—the uplink L1/L2 control signaling is time multiplexed with the coded UL-SCH on to the PUSCH prior to DFT precoding and OFDM modulation. As the terminal has been assigned UL-SCH resources, there is no need to support transmission of the scheduling request in this case. Instead, scheduling information can be included in the MAC headers, as described in Chapter 12.

Copyright Elsevier 2024

The reason to differentiate between the two cases above is to minimize the cubic metric for the uplink power amplifier in order to maximize coverage. However, in situations when there is sufficient power available in the terminal, simultaneous transmission of PUSCH and PUCCH can be used with no impact on the coverage. The possibility for simultaneous PUSCH and PUCCH transmission was therefore introduced in release 10 as one part of several features[15] adding flexibility at the cost of a somewhat higher cubic metric. In situations where this cost is not acceptable, simultaneous PUSCH and PUCCH can always be avoided by using the basic mechanism introduced in the first version of LTE.

In the following section, the basic PUCCH structure and the principles for PUCCH control signaling are described, followed by control signaling on PUSCH.

11.4.1 Basic PUCCH structure

If the terminal has not been assigned an uplink resource for UL-SCH transmission, the L1/L2 control information (CSI reports, hybrid-ARQ acknowledgements, and scheduling requests) is transmitted on uplink resources (resource blocks) specifically assigned for uplink L1/L2 control on PUCCH. The bandwidth of one resource block during one subframe is too large for the control signaling needs of a single terminal. Therefore, to efficiently exploit the resources set aside for control signaling, multiple terminals can share the same resource-block pair.

There is one PUCCH per terminal. As mentioned previously, the PUCCH is transmitted on the primary component carrier only. However, as the primary component carrier is specified on a per-terminal basis, there may, from a network perspective, be PUCCH resources used on multiple component carriers.

Transmission of control signaling on PUCCH is characterized by the PUCCH format used. There are three different PUCCH formats provided, differentiated primarily by the size of the supported payload:

- PUCCH format 1, carrying 0, 1 or 2 bits of information and used for hybrid-ARQ acknowledgements and scheduling requests.
- PUCCH format 2, carrying up to 11 bits of control information and used for reporting channel-state information.
- PUCCH format 3, carrying up to 22 bits of control information and used, among other things, for multiple hybrid-ARQ acknowledgements. This format was introduced in conjunction with carrier aggregation in release 10.

The detailed structure of each of the different PUCCH formats will be discussed below, followed by an overview on how the formats are used in the next section.

[15]Other examples of such features are simultaneous transmission on multiple uplink component carriers and uplink multi-cluster transmission.

Copyright Elsevier 2024

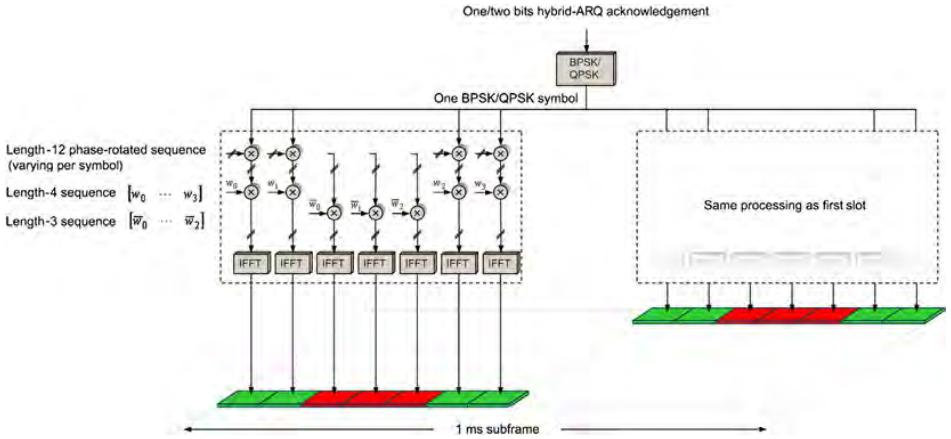


FIGURE 11.22

PUCCH format 1 (normal cyclic prefix)

11.4.1.1 PUCCH format 1

PUCCH format 1,[16] for transmission of hybrid-ARQ acknowledgements and scheduling requests, is capable of carrying up to two bits of information. The same structure is used in the two slots of a subframe as illustrated in Figure 11.22. For transmission of a hybrid-ARQ acknowledgement, the one or two hybrid-ARQ acknowledgement bits are used to generate a BPSK or QPSK, respectively. For a scheduling request, the same constellation point as for a negative acknowledgement is used. The modulation symbol is then used to generate the signal to be transmitted in each of the two PUCCH slots.

Different terminals sharing the same resource-block pair in a subframe are separated by different orthogonal phase rotations of a length-12 frequency-domain sequence, where the sequence is identical to a length-12 reference-signal sequence. Furthermore, as described in conjunction with the reference signals in Section 11.2, a linear phase rotation in the frequency domain is equivalent to applying a cyclic shift in the time domain. Thus, although the term "phase rotation" is used herein, the term cyclic shift is sometimes used with an implicit reference to the time domain. Similarly to the case of reference signals, there are up to 12 different phase rotations specified, providing up to 12 different orthogonal sequences from each base sequence[17]. However, in the case of frequency-selective channels, not all 12 phase rotations can be used if orthogonality is to be retained.

[16]There are actually three variants in the LTE specifications, formats 1, 1a, and 1b, used for transmission of scheduling requests and one or two hybrid-ARQ acknowledgements respectively. However, for simplicity, they are all referred to as format 1 herein.

[17]In releases 8 to 10, the base sequence is cell-specific, while release 11 adds the possibility of configuring a "virtual cell identity" from which the base sequence is derived. See further the discussion in conjunction with uplink reference signals in Section 11.2.

Copyright Elsevier 2024

Typically, up to six rotations are considered usable in a cell from a radio-propagation perspective, although inter-cell interference may result in a smaller number being useful from an overall system perspective. Higher-layer signaling is used to configure the number of rotations that are used in a cell.

There are seven OFDM symbols per slot for a normal cyclic prefix (six in the case of an extended cyclic prefix). In each of those seven OFDM symbols, a length-12 sequence, obtained by phase rotation of the base sequence as described earlier, is transmitted. Three of the symbols are used as reference signals to enable channel estimation by the eNodeB and the remaining four[18] are modulated by the BPSK/QPSK symbols described earlier. In principle, the BPSK/QPSK modulation symbol could directly modulate the rotated length-12 sequence used to differentiate terminals transmitting on the same time–frequency resource. However, this would result in unnecessarily low capacity on the PUCCH. Therefore, the BPSK/QPSK symbol is multiplied by a length-4 orthogonal cover sequence.[19] Multiple terminals may transmit on the same time–frequency resource using the same phase-rotated sequence and be separated through different orthogonal covers. To be able to estimate the channels for the respective terminals, the reference signals also employ an orthogonal cover sequence, with the only difference being the length of the sequence—three for the case of a normal cyclic prefix. Thus, since each base sequence can be used for up to $3 \cdot 12 = 36$ different terminals (assuming all 12 rotations are available; typically at most six of them are used), there is a threefold improvement in the PUCCH capacity compared to the case of no cover sequence. The cover sequences are three Walsh sequences of length 4 for the data part and three DFT sequences of length 3 for the reference signals.

A PUCCH format 1 resource, used for either a hybrid-ARQ acknowledgement or a scheduling request, is represented by a single scalar resource index. From the index, the phase rotation and the orthogonal cover sequence are derived.

The use of a phase rotation of a base sequence together with orthogonal sequences as described earlier provides orthogonality between different terminals in the same cell transmitting PUCCH on the same set of resource blocks. Hence, in the ideal case, there will be no intra-cell interference, which helps improve the performance. However, there will typically be inter-cell interference for the PUCCH as the different sequences used in neighboring cells are non-orthogonal. To randomize the inter-cell interference, the phase rotation of the sequence used in a cell varies on a symbol-by-symbol basis in a slot according to a hopping pattern derived from the physical-layer cell identity of the primary carrier. In release 11, the randomization can be configured to use the virtual cell identity instead of the physical one. On top of this, slot-level hopping is applied to the orthogonal cover and phase rotation to further randomize the interference. This is exemplified in Figure 11.23, assuming normal cyclic prefix

[18]The number of symbols used for reference signals and the acknowledgement is a trade-off between channel-estimation accuracy and energy in the information part; three symbols for reference symbols and four symbols for the acknowledgement have been found to be a good compromise.

[19]In the case of simultaneous SRS and PUCCH transmissions in the same subframe, a length-3 sequence is used, thereby making the last OFDM symbol in the subframe available for the sounding reference signal.

Copyright Elsevier 2024

Phase rotation [multiples of 2π/12]	Number of cover sequence					
	Even-numbered slot			Odd-numbered slot		
	0	1	2	0	1	2
0	0		12	12		16
1		6			14	
2	1		13	6		10
3		7			8	
4	2		14	0		4
5		8			2	
6	3		15	13		17
7		9			15	
8	4		16	7		11
9		10			9	
10	5		17	1		5
11		11			3	

Cell A

Phase rotation [multiples of 2π/12]	Number of cover sequence					
	Even-numbered slot			Odd-numbered slot		
	0	1	2	0	1	2
0		11			3	
1	0		12	12		16
2		6			14	
3	1		13	6		10
4		7			8	
5	2		14	0		4
6		8			2	
7	3		15	13		17
8		9			15	
9	4		16	7		11
10		10			9	
11	5		17	1		5

Cell B

FIGURE 11.23

Example of phase rotation and cover hopping for two PUCCH resource indices in two different cells

and six of 12 rotations used for each cover sequence. To the phase rotation given by the cell-specific hopping a slot-specific offset is added. In cell A, a terminal is transmitting on PUCCH resource number 3, which in this example corresponds to using the (phase rotation, cover sequence) combination (6, 0) in the first slot and (11, 1) in the second slot of this particular subframe. PUCCH resource number 11, used by another terminal in cell A transmitting in the same subframe, corresponds to (11, 1) and (8, 2) in the first and second slots respectively of the subframe. In another cell the PUCCH resource numbers are mapped to different sets (rotation, cover sequence) in the slots. This helps to randomize the inter-cell interference.

For an extended cyclic prefix, the same structure as in Figure 11.19 is used with the difference being the number of reference symbols in each slot. In this case, the six OFDM symbols in each slot are divided such that the two middle symbols are used for reference signals and the remaining four symbols used for the information. Thus, the length of the orthogonal sequence used to spread the reference symbols is reduced from 3 to 2 and the multiplexing capacity is lower. However, the general principles described above still apply.

11.4.1.2 *PUCCH format 2*

PUCCH format 2, used for CSI reporting, is capable of handling up to 11 information bits per subframe.[20] Similarly to PUCCH format 1, multiple terminals using the same resource-block

[20]There are actually three variants in the LTE specifications, formats 2, 2a, and 2b, where the last two formats are used for simultaneous transmission of CSI reports and hybrid-ARQ acknowledgements, as discussed later in this section. However, for simplicity, they are all referred to as format 2 herein.

Copyright Elsevier 2024

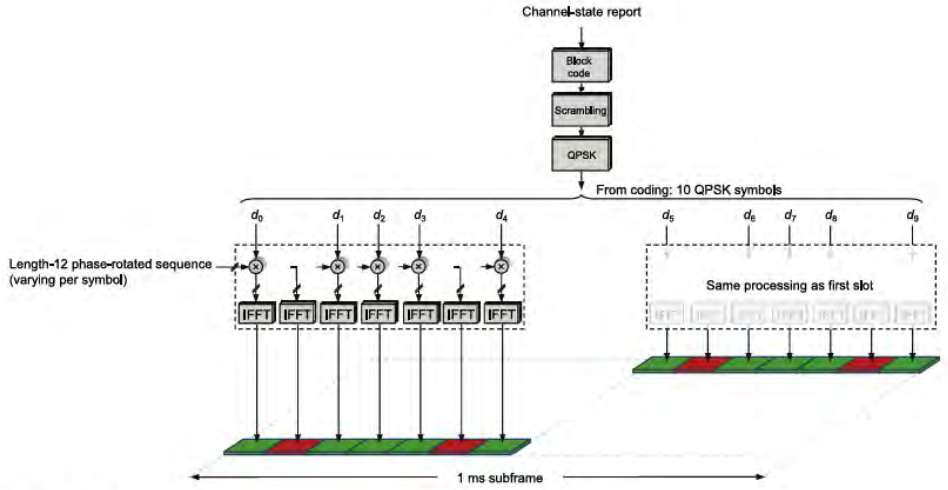


FIGURE 11.24

PUCCH format 2 (normal cyclic prefix)

pair in a subframe are separated through different orthogonal phase rotations of a length-12 sequence, as illustrated for a normal cyclic prefix in Figure 11.24. After block coding using a punctured Reed–Müller code and QPSK modulation, there are 10 QPSK symbols to transmit in the subframe: The first five symbols are transmitted in the first slot and the remaining five in the last slot.

Assuming a normal cyclic prefix, there are seven OFDM symbols per slot. Of the seven OFDM symbols in each slot, two[21] are used for reference-signal transmission to allow coherent demodulation at the eNodeB. In the remaining five, the respective QPSK symbol to be transmitted is multiplied by a phase-rotated length-12 base sequence and the result is transmitted in the corresponding OFDM symbol. For an extended cyclic prefix, where there are six OFDM symbols per slot, the same structure is used but with one reference-signal symbol per slot instead of two.

Basing the format 2 structure on phase rotations of the same base sequence as format 1 is beneficial as it allows the two formats to be transmitted in the same resource block. As phase-rotated sequences are orthogonal, one rotated sequence in the cell can be used either for one

[21]Similarly to format 1, the number of symbols used for reference signals and the coded channel-quality information is a trade-off between channel-estimation accuracy and energy in the information part. Two symbols for reference symbols and five symbols for the coded information part in each slot were found to be the best compromise.

Copyright Elsevier 2024

Samsung Exhibit 1017, Page 298 of 538
Samsung Electronics Co., Ltd. v. Asus Technology Licensing Inc.
IPR2024-01217

PUCCH instance using format 2 or three PUCCH instances using format 1. Thus, the "resource consumption" of one CSI report is equivalent to three hybrid-ARQ acknowledgements (assuming normal cyclic prefix). Note that no orthogonal cover sequences are used for format 2.

The phase rotations to use in the different symbols for PUCCH format 2 are hopping in a similar way as for format 1, motivated by interference randomization. Resources for PUCCH format 2 can, similar to format 1, be represented by a scalar index, which can be seen as a "channel number."

11.4.1.3 PUCCH format 3

For downlink carrier aggregation, multiple hybrid-ARQ acknowledgement bits need to be fed back in the case of simultaneous transmission on multiple component carriers. Although PUCCH format 1 with resource selection can be used to handle the case of two downlink component carriers, this is not sufficient as a general solution as carrier aggregation of up to five component carriers is part of release 10. PUCCH format 3 was therefore introduced in release 10 to enable the possibility of transmitting up to 22 bits on PUCCH in an efficient way. A terminal capable of more than two downlink component carriers—that is, capable of more than four bits for hybrid-ARQ acknowledgements—needs to support PUCCH format 3. For such a terminal, PUCCH format 3 can also be used for less than four bits of feedback relating to simultaneous transmission on multiple component carriers if configured by higher-layer signaling not to use PUCCH format 1 with resource selection.

The basis for PUCCH format 3, illustrated in Figure 11.25, is DFT-precoded OFDM—that is, the same transmission scheme as used for UL-SCH. The acknowledgement bits, one or two per downlink component carrier depending on the transmission mode configured for that particular component carrier, are concatenated with a bit reserved for scheduling a request into a sequence of bits where bits corresponding to unscheduled transport blocks are set to zero. Block coding is applied,[22] followed by scrambling to randomize inter-cell interference. The resulting 48 bits are QPSK-modulated and divided into two groups, one per slot, of 12 QPSK symbols each.

Assuming a normal cyclic prefix, there are seven OFDM symbols per slot. Similarly to PUCCH format 2, two OFDM symbols (one in the case of an extended cyclic prefix) in each slot are used for reference signal transmission, leaving five symbols for data transmission. In each slot, the block of 12 DFT-precoded QPSK symbols is transmitted in the five available DFTS-OFDM symbols. To further randomize the inter-cell interference, a cyclic shift of the 12 inputs to the DFT, varying between OFDM symbols in a cell-specific manner, is applied to the block of 12 QPSK symbols prior to DFT precoding (in release 11, the cyclic shift can be based on the virtual cell identity instead of the physical one).

[22] A (32,k) Reed-Müller code is used, but for 12 or more bits in TDD two Reed-Müller codes are used in combination.

Copyright Elsevier 2024

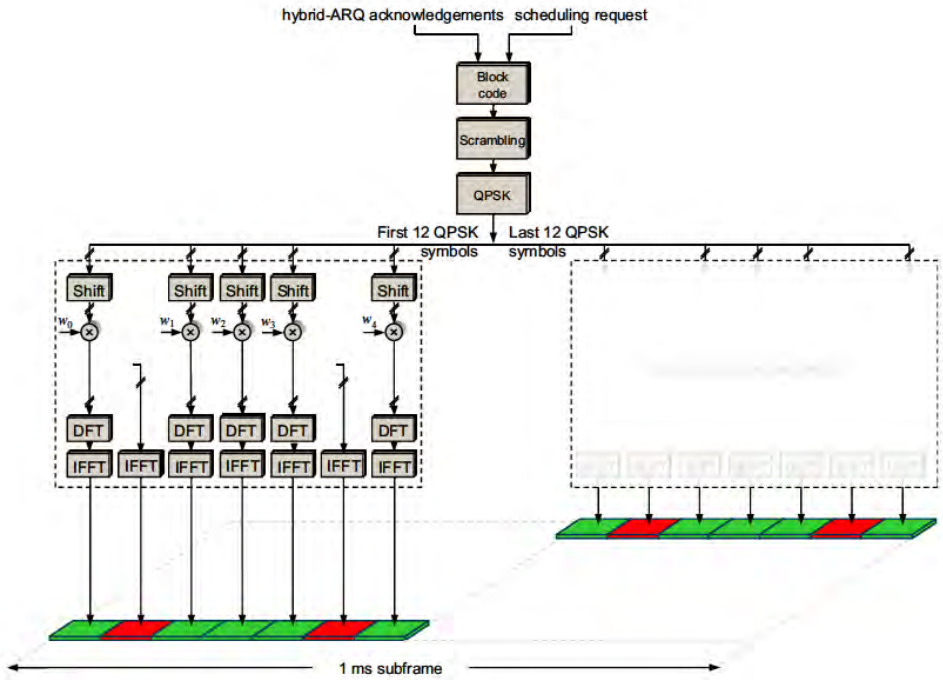


FIGURE 11.25

PUCCH format 3 (normal cyclic prefix)

To increase the multiplexing capacity, a length-5 orthogonal sequence is used with each of the five OFDM symbols carrying data in a slot being multiplied by one element of the sequence. Thus, up to five terminals may share the same resource-block pair for PUCCH format 3. Different length-5 sequences are used in the two slots to improve the performance in high-Doppler scenarios. To facilitate channel estimation for the different transmissions sharing the same resource block, different reference-signal sequences are used.

The length-5 orthogonal cover sequences are obtained as five DFT sequences. There is also the possibility to use a length-4 Walsh sequence for the second slot in order to leave the last OFDM symbol unused for the case when sounding is configured in the subframe.

In the same manner as for the other two PUCCH formats, a resource can be represented by a single index from which the orthogonal sequence and the resource-block number can be derived.

Note that, due to the differences in the underlying structure of PUCCH format 3 compared to the other two formats, resource blocks cannot be shared between format 3 and the other two formats.

Copyright Elsevier 2024

11.4.1.4 Resource-block mapping for PUCCH

The signals described for all three of the PUCCH formats are, as already explained, transmitted on a resource-block pair. The resource-block pair to use is determined from the PUCCH resource index. Multiple resource-block pairs can be used to increase the control-signaling capacity in the cell; when one resource-block pair is full, the next PUCCH resource index is mapped to the next resource-block pair in sequence.

The resource-block pair where a PUCCH is transmitted is located at the edges of the bandwidth allocated to the primary component carrier,[23] as illustrated in Figure 11.26. To provide frequency diversity, frequency hopping on the slot boundary is used—that is, one "frequency resource" consists of 12 subcarriers at the upper part of the spectrum within the first slot of a subframe and an equally sized resource at the lower part of the spectrum during the second slot of the subframe (or vice versa).

The reasons for locating the PUCCH resources at the edges of the overall available spectrum are twofold:

- Together with the frequency hopping described previously, this maximizes the frequency diversity experienced by the control signaling.
- Assigning uplink resources for the PUCCH at other positions within the spectrum—that is, not at the edges—would have fragmented the uplink spectrum, making it impossible to assign very wide transmission bandwidths to a single terminal and still preserve the low-cubic-metric properties of the uplink transmission.

The resource-block mapping is in principle done such that PUCCH format 2 (CSI reports) is transmitted closest to the edges of the uplink cell bandwidth with PUCCH

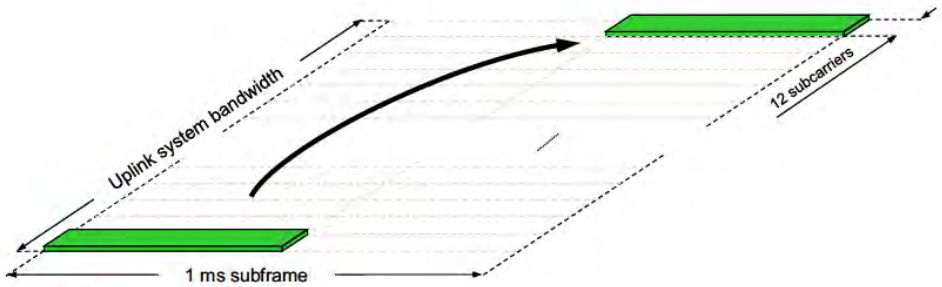


FIGURE 11.26

Uplink L1/L2 control signaling transmission on PUCCH

[23]Note that the primary component carrier in the uplink is specified on a per-terminal basis. Hence, different terminals may view different carriers as their primary component carrier.

Copyright Elsevier 2024

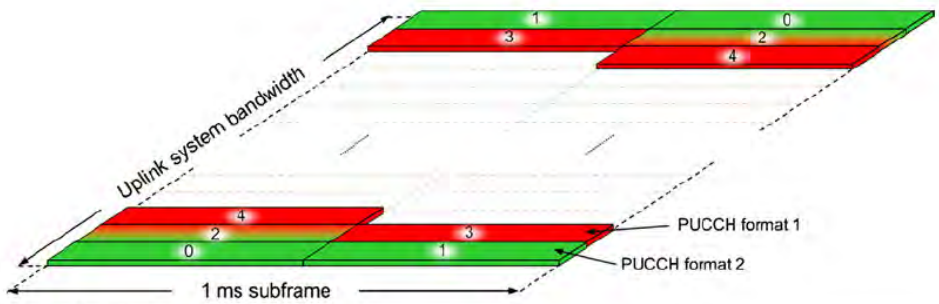


FIGURE 11.27

Allocation of resource blocks for PUCCH

format 1 (hybrid-ARQ acknowledgements, scheduling requests) next, as illustrated in Figure 11.27. The location of PUCCH format 3 is configurable and it can, for example, be located between formats 1 and 2. A semi-static parameter, provided as part of the system information, controls on which resource-block pair the mapping of PUCCH format 1 starts. Furthermore, the semi-statically configured scheduling requests are located at the outermost parts of the format 1 resources, leaving dynamic acknowledgements closest to the data. As the amount of resources necessary for hybrid-ARQ acknowledgements varies dynamically, this maximizes the amount of contiguous spectrum available for PUSCH.

In many scenarios, the configuration of the PUCCH resources can be done such that the three PUCCH formats are transmitted on separate sets of resource blocks. However, for the smallest cell bandwidths, this would result in too high an overhead. Therefore, it is

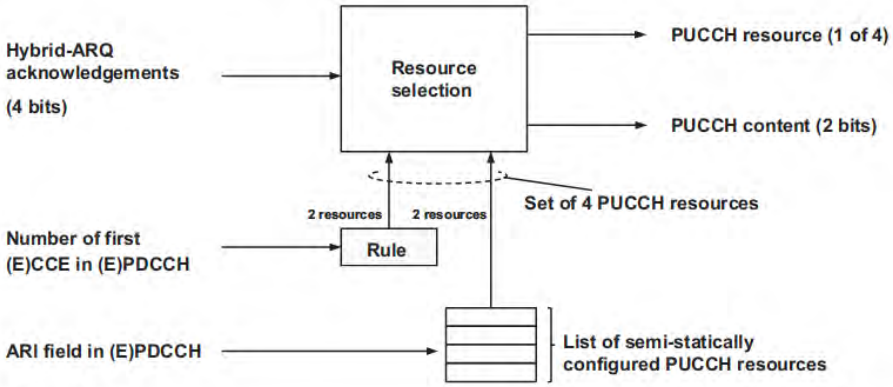


FIGURE 11.28

Example illustration of resource selection using PUCCH format 1 for carrier aggregation

Copyright Elsevier 2024

Table 11.3 Usage of Different PUCCH Formats for Different Pieces of Information (the Superscripts 10 and 11 Denote the First Release Supporting this Combination)

	PUCCH Format				
	Format 1				
Information		Selection	Bundling	Format 2	Format 3
ACK	•	•[10]			•[10]
SR	•				
SR + ACK	•		•[10]		•[10]
CSI				•	
CSI + ACK				•	•[11]
CSI + SR					
CSI + SR + ACK					•[11]

possible to mix PUCCH formats 1 and 2 in one of the resource-block pairs—for example, in Figure 11.27 this is the case for the resource-block pair denoted "2." Although this mixture is primarily motivated by the smaller cell bandwidths, it can equally well be used for the larger cell bandwidths. In the resource-block pair where PUCCH formats 1 and 2 are mixed, the set of possible phase rotations are split between the two formats. Furthermore, some of the phase rotations are reserved as "guard," hence the efficiency of such a mixed resource-block pair is slightly lower than a resource-block pair carrying only one of the first two PUCCH formats.

11.4.2 Uplink control signaling on PUCCH

Having described the three PUCCH formats, the details on how these different formats are used to convey uplink control information can be discussed. As already mentioned, uplink control signaling on PUCCH can in principle be any combination of hybrid-ARQ acknowledgements (ACK), channel-state information (CSI), and scheduling requests (SR). Depending on whether these pieces of information are transmitted alone or in combination, different PUCCH formats and mechanisms are used, as summarized in Table 11.3. In principle, simultaneous transmission of multiple control signaling messages from a single terminal could use multiple PUCCHs. However, this would increase the cubic metric, and a single PUCCH structure supporting simultaneous transmission of multiple feedback signals is used instead.

11.4.2.1 Hybrid-ARQ acknowledgements

Hybrid-ARQ acknowledgements are used to acknowledge receipt of one (or two in the case of spatial multiplexing) transport blocks on the DL-SCH. PUCCH format 1 is used in absence of carrier aggregation but can also support carrier aggregation of up to two downlink

Copyright Elsevier 2024

carriers—that is, up to four acknowledgement bits—as discussed below. PUCCH format 3 is used for more than four acknowledgements bits.

The hybrid-ARQ acknowledgement is only transmitted when the terminal correctly received control signaling related to DL-SCH transmission intended for this terminal on an PDCCH or EPDCCH. If no valid DL-SCH-related control signaling is detected, then nothing is transmitted on the PUCCH (that is, DTX). Apart from not unnecessarily occupying PUCCH resources that can be used for other purposes, this allows the eNodeB to perform three-state detection, ACK, NAK, or DTX, on the PUCCH received when using PUCCH format 1. Three-state detection is useful as NAK and DTX may need to be treated differently. In the case of NAK, retransmission of additional parity bits is useful for incremental redundancy, while for DTX the terminal has most likely missed the initial transmission of systematic bits and a better alternative than transmitting additional parity bits is to retransmit the systematic bits.

Transmission of one or two hybrid-ARQ acknowledgement bits uses PUCCH format 1. As mentioned in Section 11.4.1.1, a PUCCH resource can be represented by an index. How to determine this index depends on the type of information and whether the PDCCH or the EPDCCH was used to schedule the downlink data transmission.

For PDCCH-scheduled downlink transmissions, the resource index to use for a hybrid-ARQ acknowledgement is given as a function of the first CCE in the PDCCH used to schedule the downlink transmission to the terminal. In this way, there is no need to explicitly include information about the PUCCH resources in the downlink scheduling assignment, which of course reduces overhead. Furthermore, as described in Chapter 12, hybrid-ARQ acknowledgements are transmitted a fixed time after the reception of a DL-SCH transport block, and when to expect a hybrid ARQ on the PUCCH is therefore known to the eNodeB.

For EPDCCH-scheduled transmissions, the index of the first ECCE in the EPDCCH cannot be used alone. Since the ECCE numbering is configured per terminal and therefore is terminal-specific, two different terminals with control signaling on different resource blocks may have the same number of the first ECCE in the EPDCCH. Therefore, the ACK/NAK resource offset (ARO) being part of the EPDCCH information (see Section 10.4.6) is used in addition to the index of the first ECCE to determine the PUCCH resource. In this way, PUCCH collisions between two terminals scheduled with EPDCCH can be avoided.

In addition to dynamic scheduling by using the (E)PDCCH, there is also, as described in Chapter 13, the possibility to semi-persistently schedule a terminal according to a specific pattern. In this case there is no PDCCH or EPDCCH to derive the PUCCH resource index from. Instead, the configuration of the semi-persistent scheduling pattern includes information on the PUCCH index to use for the hybrid-ARQ acknowledgement. In either of these cases, a terminal is using PUCCH resources only when it has been scheduled in the downlink. Thus, the amount of PUCCH resources required for hybrid-ARQ acknowledgements does not necessarily increase with an increasing number of terminals in the cell, but, for dynamic scheduling, is rather related to the number of CCEs in the downlink control signaling.

Copyright Elsevier 2024

The description above addressed the case of downlink carrier aggregation not being used. For downlink carrier aggregation, introduced in release 10, there can be multiple simultaneous DL-SCHs scheduled for a single terminal, one per downlink component carrier, and, consequently, multiple acknowledgement bits need to be conveyed in the uplink (one, or two in the case of spatial multiplexing, per downlink component carrier). PUCCH format 1 can be used to support more than two bits in the uplink by using resource selection where parts of the information are conveyed by the PUCCH resource selected, and parts by the bits transmitted on the selected resource. The details on how to select the resources are rather complex and depend on the duplexing scheme (FDD or TDD) and whether cross-carrier scheduling is used or not, although the basic idea is the same.

As an example, assume FDD with no cross-carrier scheduling. Furthermore, assume four bits are to be transmitted in the uplink—that is, there are 16 possible combinations of positive and negative acknowledgements. For each of these 16 combinations, one PUCCH resource out of four possible resources is selected and upon this resource two bits are transmitted. Thus, to convey a total of four bits, there are four candidate PUCCH resources for each of the 16 combinations.

Two of the PUCCH candidate resources to select from are derived from the first (E)CCE using the same rule as in absence of carrier aggregation (assuming that the scheduling assignment is transmitted on, and relating to, the primary component carrier). The remaining two candidate resources are obtained by the *acknowledgement resource indicator* (ARI, see Section 10.4.6) on the (E)PDCCH pointing to a list of semi-statically configured PUCCH resources. For transmissions scheduled by the EPDCCH, the *acknowledgement resource offset* (ARO) is also included in the determination of the PUCCH resources. By setting the ARO appropriately, the scheduler can ensure that multiple terminals use non-colliding PUCCH resources. In presence of cross-carrier scheduling, all four resources are semi-statically configured, and the ARI is not used.

For more than four bits, resource selection is less efficient and PUCCH format 3, introduced in release 10, is used. In PUCCH format 3, the set of multiple hybrid-ARQ acknowledgement bits (up to 10 or 20 bits for FDD or TDD, respectively, plus one bit reserved for scheduling requests) are jointly coded and transmitted as described in Section 11.4.1.3. Not all terminals support PUCCH format 3, but for those that do support it, it can be used also for smaller numbers of acknowledgement bits. The PUCCH resource to use is determined by the ARI. A terminal can be configured with four different resources for PUCCH format 3 using RRC signaling. In the scheduling grant for a secondary carrier, the ARI informs the terminal which of the four resources to use. In this way, the scheduler can avoid PUCCH collisions between different terminals by assigning them to different resources.

11.4.2.2 Scheduling request

Scheduling requests are used to request uplink resources for data transmission. A scheduling request should only be transmitted when the terminal is requesting resources;

Copyright Elsevier 2024

otherwise, the terminal should be silent to save battery resources and not create unnecessary interference.

Unlike the hybrid-ARQ acknowledgements, whose occurrence is known to the eNodeB from the downlink scheduling decisions, the need for uplink resources for a certain terminal is in principle unpredictable by the eNodeB. One way to handle this would be to have a contention-based mechanism for requesting uplink resources. The random-access mechanism is based on this principle and can, to some extent, also be used for scheduling requests, as discussed in Chapter 13. Contention-based mechanisms typically work well for low intensities, but for higher scheduling-request intensities, the collision rate between different terminals simultaneously requesting resources becomes too large. Therefore, LTE provides a contention-free scheduling-request mechanism on the PUCCH, where each terminal in the cell is given a reserved resource on which it can transmit a request for uplink resources. Unlike hybrid-ARQ acknowledgements, no explicit information bit is transmitted by the scheduling request; instead the information is conveyed by the presence (or absence) of energy on the corresponding PUCCH. However, the scheduling request, although used for a completely different purpose, shares the same PUCCH format as the hybrid-ARQ acknowledgement, namely PUCCH format 1.

The contention-free scheduling-request resource is represented by a PUCCH format 1 resource index as described earlier, occurring at every *n*th subframe. The more frequently these time instants occur, the lower the scheduling-request delay at the cost of higher PUCCH resource consumption. As the eNodeB configures all the terminals in the cell, when and on which resources a terminal can request resources is known to the eNodeB. A single scheduling request resource is also sufficient for the case of carrier aggregation, as it only represents a request for uplink resources, which is independent of whether carrier aggregation is used or not.

11.4.2.3 Hybrid-ARQ acknowledgements and scheduling request

The discussion in the previous two sections concerned transmission of *either* a hybrid-ARQ acknowledgement *or* a scheduling request. However, there are situations when the terminal needs to transmit *both* of them.

If PUCCH format 1 is used for the acknowledgements, simultaneous transmission of the acknowledgements and scheduling request is handled by transmitting the hybrid-ARQ acknowledgement on the scheduling-request resource (see Figure 11.29). This is possible



FIGURE 11.29

Multiplexing of scheduling request and hybrid-ARQ acknowledgement from a single terminal

Copyright Elsevier 2024

as the same PUCCH structure is used for both of them and the scheduling request carries no explicit information. By comparing the amount of energy detected on the acknowledgement resource and the scheduling-request resource for a specific terminal, the eNodeB can determine whether the terminal is requesting uplink data resources or not. Once the PUCCH resource used for transmission of the acknowledgement is detected, the hybrid-ARQ acknowledgement can be decoded. Other, more advanced methods jointly decoding hybrid ARQ and scheduling request can also be envisioned.

Channel selection, which is a way to transmit up to four acknowledgements on PUCCH in absence of a simultaneous scheduling request, cannot be used for joint transmission of acknowledgements and scheduling request. Instead, up to four acknowledgement bits are bundled (combined) into two bits that are transmitted, as described in the previous paragraph. Bundling implies that two or more acknowledgement bits are combined into a smaller number of bits. In essence, one acknowledgement bit represents the decoding outcome of multiple transport blocks, and all these transport blocks need to be retransmitted as soon as one of them is incorrectly received.

If the hybrid-ARQ acknowledgement is transmitted using PUCCH format 3, the scheduling request is encoded jointly with the acknowledgement bits by setting the bit reserved for scheduling requests in PUCCH format 3 to one. No bundling is needed in this case.

11.4.2.4 Channel-state information

CSI reports, the contents of which are discussed in Chapter 13, are used to provide the eNodeB with an estimate of the downlink radio-channel properties as seen from the terminal to aid channel-dependent scheduling. A CSI report consists of multiple bits transmitted in one subframe. There are *two* types of CSI reports:

- *Periodic* reports, occurring at regular time instants
- *Aperiodic* reports, triggered by downlink control signaling on the PDCCH or EPDCCH

Aperiodic reports can only be transmitted on PUSCH as described later in Section 11.4.3, while periodic reports can be transmitted on PUCCH using PUCCH format 2.

A PUCCH format 2 resource is represented by an index and higher-layer signaling is used to configure each terminal with a resource to transmit its CSI report on, as well as when those reports should be transmitted. Hence, the eNodeB has full knowledge of when and on which resources each of the terminals will transmit channel-state information on PUCCH.

Periodic CSI reports in combination with carrier aggregation are provided for by transmitting multiple periodic reports, one per component carrier, offset in time such that reports for two different component carriers do not collide.

11.4.2.5 Hybrid-ARQ acknowledgements and channel-state information

Transmission of data in the downlink implies transmission of hybrid-ARQ acknowledgements in the uplink. At the same time, since data is transmitted in the downlink, up-to-date

Copyright Elsevier 2024

channel-state information is beneficial to optimize the downlink transmissions. Hence, simultaneous transmission of hybrid-ARQ acknowledgements and channel-state information is supported by LTE.

The handling of simultaneous transmission of acknowledgements and a CSI report depends on the number of acknowledgement bits as well as higher-layer configuration. There is also the possibility to configure the terminal to drop the CSI report and only transmit the acknowledgements.

The basic way of supporting transmission of one or two acknowledgements simultaneously with channel-state information, part of releases 8 and later, is based on PUCCH format 2, although the detailed solution differs between the two.

For a normal cyclic prefix, each slot in PUCCH format 2 has two OFDM symbols used for reference signals. When transmitting a hybrid-ARQ acknowledgement at the same time as the CSI report, the second reference signal in each slot is modulated by the acknowledgement, as illustrated in Figure 11.30a. Either BPSK or QPSK is used, depending on whether one or two acknowledgement bits are to be fed back. The fact that the acknowledgement is superimposed on the reference signal needs to be accounted for at the eNodeB. One possibility is to decode the acknowledgement bit(s) modulated on to the second reference symbol using the first reference symbol for channel estimation. Once the acknowledgement bit(s) have been decoded, the modulation imposed on the second reference symbol can be removed and channel estimation and decoding of the CSI report can be handled in the same way as in the absence of simultaneous hybrid-ARQ acknowledgement. This two-step approach works well for low to medium Doppler frequencies; for higher Doppler frequencies, the acknowledgement and CSI reports are preferably decoded jointly.

For an extended cyclic prefix, there is only a single reference symbol per slot. Hence, it is not possible to overlay the hybrid-ARQ acknowledgement on the reference symbol. Instead,

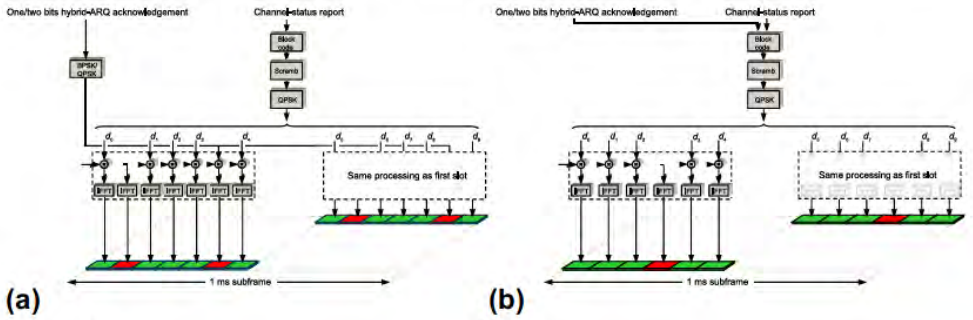


FIGURE 11.30

Simultaneous transmission of channel-state information and hybrid-ARQ acknowledgements: (a) normal cyclic prefix and (b) extended cyclic prefix

Copyright Elsevier 2024

Samsung Exhibit 1017, Page 308 of 538
Samsung Electronics Co., Ltd. v. Asus Technology Licensing Inc.
IPR2024-01217

the acknowledgement bit(s) are jointly coded with the CSI report prior to transmission using PUCCH format 2, as illustrated in Figure 11.30b.

The time instances for which to expect CSI reports and hybrid-ARQ acknowledgements are known to the eNodeB, which therefore knows whether to expect a hybrid-ARQ acknowledgement along with the CSI report or not. If the (E)PDCCH assignment is missed by the terminal, then only the CSI report will be transmitted, as the terminal is not aware that it has been scheduled. In the absence of a simultaneous CSI report, the eNodeB can employ DTX detection to discriminate between a missed assignment and a failed decoding of downlink data. However, one consequence of the structures described above is that DTX detection is cumbersome, if not impossible. This implies that incremental redundancy needs to be operated with some care if the eNodeB has scheduled data such that the acknowledgement occurs at the same time as a CSI report. As the terminal may have missed the original transmission attempt in the downlink, it may be preferable for the eNodeB to select the redundancy version of the retransmission such that systematic bits are also included in the retransmission.

One possibility to circumvent this is to configure the terminal to drop the CSI report in the case of simultaneous transmission of a hybrid-ARQ acknowledgement. In this case, the eNodeB can detect DTX as the acknowledgement can be transmitted using PUCCH format 1 as described earlier. There will be no CSI report sent in this case, which needs to be taken into account in the scheduling process.

For a larger number of acknowledgements in case of carrier aggregation, PUCCH format 3 can be used to transmit the acknowledgements. In release 10, simultaneous transmission of channel-state information and acknowledgements on PUCCH format 3 is not supported and the CSI report is dropped. Release 11, on the other hand, provides support for simultaneous transmission of acknowledgements and periodic CSI reports. The CSI bits are in this case concatenated with the acknowledgement bits and the scheduling request bit and transmitted using PUCCH format 3. The PUCCH resource to use is determined by the ARI in the downlink control signaling, which selects the resource to use from a set of four resources preconfigured in the terminal by RRC signaling.

If the number of concatenated bits is larger than 22 bits (the maximum payload of PUCCH format 3), bundling of the acknowledgement bits are applied prior to concatenation in order to reduce the payload size. In case the payload still is larger than 22 bits, then the CSI is dropped and the acknowledgements (and potentially scheduling request) are transmitted as previously described. The process of bundling and dropping in case of a too large payload is illustrated in Figure 11.31.

11.4.2.6 Scheduling request and channel-state information

The eNodeB is in control of when a terminal may transmit a scheduling request and when the terminal should report the channel state. Hence, simultaneous transmission of scheduling requests and channel-state information can be avoided by proper configuration. If this is not done, the terminal drops the CSI report and transmits the scheduling request

Copyright Elsevier 2024

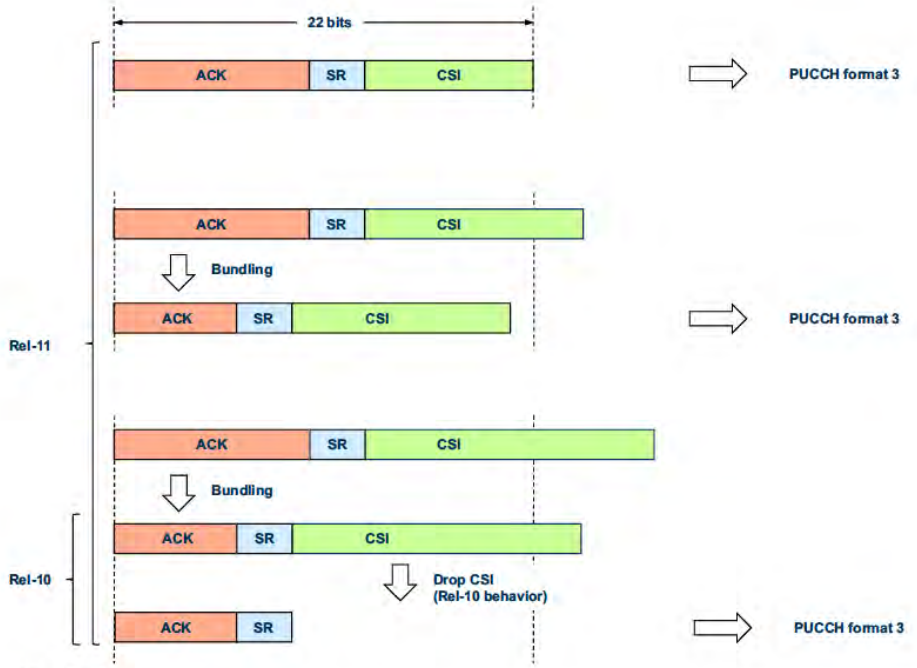


FIGURE 11.31

Multiplexing of acknowledgements, scheduling request and CSI in release 11

only. Missing a CSI report is not detrimental and only incurs some degradation in the scheduling and rate-adaptation accuracy, whereas the scheduling request is critical for uplink transmissions.

11.4.2.7 Hybrid-ARQ acknowledgements, channel-state information, and scheduling request

For terminals not supporting or configured to use PUCCH format 3, simultaneous transmission of acknowledgements, channel-state information, and scheduling request is handled similarly to the description in the previous section; the CSI report is dropped and the acknowledgements and scheduling request are multiplexed as previously described in Section 11.4.2.3. However, terminals that are using PUCCH format 3 for multiple acknowledgements support simultaneous transmission of all three pieces of information. Since there is a bit reserved for scheduling requests in PUCCH format 3, the transmission structure is no different from the case of simultaneous transmission of acknowledgements and CSI reports described in Section 11.4.2.5.

Copyright Elsevier 2024

11.4.3 Uplink L1/L2 control signaling on PUSCH

If the terminal is transmitting data on PUSCH—that is, has a valid scheduling grant in the subframe—control signaling is time multiplexed[24] with data on the PUSCH instead of using the PUCCH (in release 10, simultaneous PUSCH and PUCCH can be used, avoiding the need for control signaling on PUSCH for most cases at the cost of a somewhat worse cubic metric). Only hybrid-ARQ acknowledgements and CSI reports are transmitted on the PUSCH. There is no need to request a scheduling grant when the terminal is already scheduled; instead, in-band buffer-status reports are sent as part of the MAC headers, as described in Chapter 13.

Time multiplexing of CSI reports and hybrid-ARQ acknowledgements is illustrated in Figure 11.32. However, although they both use time multiplexing there are some differences

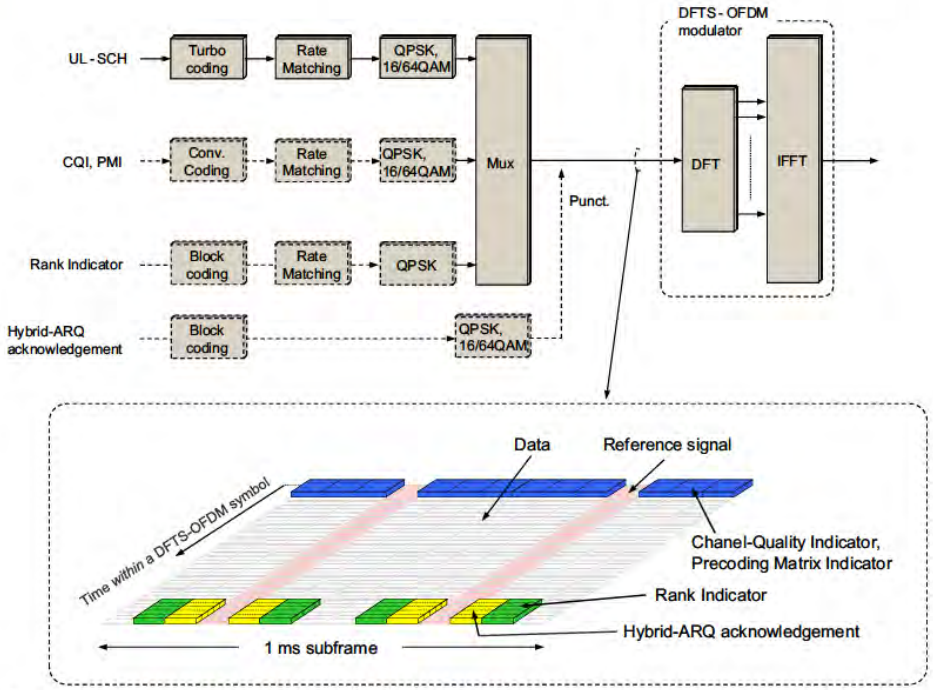


FIGURE 11.32

Multiplexing of control and data onto PUSCH

[24]In the case of spatial multiplexing, the CQI/PMI is time multiplexed with one of the codewords, implying that it is spatially multiplexed with the other codeword.

Copyright Elsevier 2024

in the details for the two types of uplink L1/L2 control signaling motivated by their different properties.

The hybrid-ARQ acknowledgement is important for proper operation of the downlink. For one and two acknowledgements, robust QPSK modulation is used, regardless of the modulation scheme used for the data, while for a larger number of bits the same modulation scheme as for the data is used. Channel coding for more than two bits is done in the same way as for the PUCCH and bundling is applied if the number of bits exceeds 20—that is, two transport blocks on the same component carrier share a single bit instead of having independent bits. Furthermore, the hybrid-ARQ acknowledgement is transmitted near to the reference symbols as the channel estimates are of better quality close to the reference symbols. This is especially important at high Doppler frequencies, where the channel may vary during a slot. Unlike the data part, the hybrid-ARQ acknowledgement cannot rely on retransmissions and strong channel coding to handle these variations.

In principle, the eNodeB knows when to expect a hybrid-ARQ acknowledgement from the terminal and can therefore perform the appropriate demultiplexing of the acknowledgement and the data part. However, there is a certain probability that the terminal has missed the scheduling assignment on the downlink control channels (PDCCH or EPDCCH), in which case the eNodeB will expect a hybrid-ARQ acknowledgement while the terminal will not transmit one. If the rate-matching pattern was to depend on whether an acknowledgement is transmitted or not, all the coded bits transmitted in the data part could be affected by a missed assignment, which is likely to cause the UL-SCH decoding to fail. To avoid this error, the hybrid-ARQ acknowledgements are therefore punctured into the coded UL-SCH bit stream. Thus, the non-punctured bits are not affected by the presence/absence of hybrid-ARQ acknowledgements and the problem of a mismatch between the rate matching in the terminal and the eNodeB is avoided.

The contents of the CSI reports are described in Chapter 13; at this stage it suffices to note that a CSI report consists of *Channel-Quality Indicator* (CQI), *Precoding Matrix Indicator* (PMI), and *Rank Indicator* (RI). The CQI and PMI are time multiplexed with the coded data bits from PUSCH and transmitted using the same modulation as the data part. CSI reports are mainly useful for low-to-medium Doppler frequencies for which the radio channel is relatively constant, hence the need for special mapping is less pronounced. The RI, however, is mapped differently than the CQI and PMI; as illustrated in Figure 11.32, the RI is located near to the reference symbols using a similar mapping as the hybrid-ARQ acknowledgements. The more robust mapping of the RI is motivated by the fact that the RI is required in order to correctly interpret the CQI/PMI. The CQI/PMI, on the other hand, is simply mapped across the full subframe duration. Modulation-wise, the RI uses QPSK.

For uplink spatial multiplexing, in which case two transport blocks are transmitted simultaneously on the PUSCH, the CQI and PMI are multiplexed with the coded transport block using the highest modulation-and-coding scheme, followed by applying the multiplexing scheme described above per layer (Figure 11.33). The intention behind

Copyright Elsevier 2024

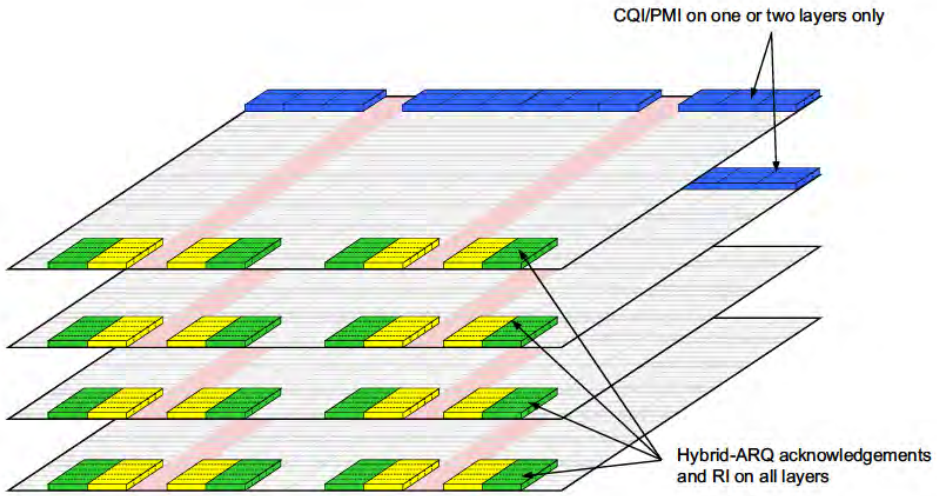


FIGURE 11.33

Multiplexing of CQI/PMI, RI and hybrid-ARQ acknowledgements in case of uplink spatial multiplexing

this approach is to transmit the CQI and PMI on the (one or two) layers with the best quality.[25]

The hybrid-ARQ acknowledgements and the rank indicator are replicated across all transmission layers and multiplexed with the coded data in each layer in the same way as the single layer case described above. The bits may, though, have been scrambled differently on the different layers. In essence, as the same information is transmitted on multiple layers with different scrambling, this provides diversity.

The basis for CSI reporting on the PUSCH is *aperiodic reports*, where the eNodeB requests a report from the terminal by setting the CSI request bit in the scheduling grant, as mentioned in Chapter 10. UL-SCH rate matching takes the presence of the CSI reports into account; by using a higher code rate a suitable number of resource elements is made available for transmission of the CSI report. Since the reports are explicitly requested by the eNodeB, their presence is known and the appropriate rate de-matching can be done at the receiver. If one of the configured transmission instances for a periodic report coincides with the terminal being scheduled on the PUSCH, the periodic report is "rerouted" and transmitted on the PUSCH resources. Also, in this case there is no risk of mismatch in rate matching;

[25] Assuming the MCS follows the channel quality, this holds for one, two, and four layers, but not necessarily for three layers.

Copyright Elsevier 2024

the transmission instants for periodic reports are configured by robust RRC signaling and the eNodeB knows in which subframes such reports will be transmitted.

The channel coding of the CSI reports depends on the report size. For the smaller sizes such as a periodic report that otherwise would have been transmitted on the PUCCH, the same block coding as used for the PUCCH reports is used. For the larger reports, a tail-biting convolutional code is used for CQI/PMI, whereas the RI uses a (3, 2) block code for a single component carrier.

Unlike the data part, which relies on rate adaptation to handle different radio conditions, this cannot be used for the L1/L2 control-signaling part. Power control could, in principle, be used as an alternative, but this would imply rapid power variations in the time domain, which negatively impact the RF properties. Therefore, the transmission power is kept constant over the subframe and the amount of resource elements allocated to L1/L2 control signaling—that is, the code rate of the control signaling—is varied according to the scheduling decision for the data of the data part. High data rates are typically scheduled when the radio conditions are advantageous and hence a smaller amount of resource needs to be used by the L1/L2 control signaling compared to the case of poor radio conditions. To account for different hybrid-ARQ operating points, an offset between the code rate for the control-signaling part and the modulation-and-coding scheme used for the data part can be configured via higher-layer signaling.

For carrier aggregation, control signaling is time multiplexed on one uplink component carrier only—that is, uplink control information cannot be split across multiple uplink component carriers. Apart from the aperiodic CSI reports, which are transmitted upon the component carrier that triggered the report, the primary component carrier is used for uplink control signaling if scheduled in the same subframe, otherwise one of the secondary component carriers is used.

11.5 Uplink power control

Uplink power control for LTE is the set of algorithms and tools by which the transmit powers for different uplink physical channels and signals are controlled to ensure that they, to the extent possible, are received with the appropriate power. This means that the transmission should be received with sufficient power to allow for proper demodulation of the corresponding information. At the same time, the transmit power should not be unnecessarily high, as that would cause unnecessary interference to other transmissions in the same or other cells. The transmit power will thus depend on the channel properties, including the channel attenuation and the noise and interference level at the receiver side. Furthermore, in the case of DL-SCH transmission on PDSCH, if the received power is too low, one can either increase the transmit power or reduce the data rate by use of rate control. Thus, for PUSCH transmission there is an intimate relation between power control and the link adaptation (rate control).

How to set the transmit power for random access will be discussed in Chapter 14. Here we will mainly discuss the power-control mechanism for the PUCCH and PUSCH physical

Copyright Elsevier 2024

channels. We will also briefly discuss the power setting for sounding reference signals (SRS). Uplink demodulation reference signals are always transmitted together and time-multiplexed with PUSCH or PUCCH. The demodulation reference signals are then transmitted with the same power as the corresponding physical channel. This is also true in the case of uplink spatial multiplexing if the reference signal power is defined as the total power of all demodulation reference signals transmitted by the terminal. Expressed differently, the power of a single demodulation reference signal is equal to the corresponding *per-layer* PUSCH power.

Fundamentally, LTE uplink power control is a combination of an *open-loop* mechanism, implying that the terminal transmit power depends on estimates of the downlink path loss, and a *closed-loop* mechanism, implying that the network can, in addition, directly adjust the terminal transmit power by means of explicit *power-control commands* transmitted on the downlink. In practice, these power-control commands are determined based on prior network measurements of the received uplink power, thus the term "*closed loop*."

11.5.1 Uplink power control: Some basic rules

Before going into the details of the power control algorithms for PUSCH and PUCCH, some basic rules for the power assignment to different physical channels will be discussed. These rules mainly deal with the presence of different transmit-power limitations and how these limitations impact the transmit-power setting for different physical channels. This is especially of interest in the case of the simultaneous transmission of multiple physical channels from the same terminal, a situation that may occur for LTE releases 10 and beyond:

- Release 10 introduced the possibility for *carrier aggregation*, implying that multiple PUSCH may be transmitted in parallel on different component carriers.
- Release 10 also introduced the possibility for *simultaneous PUSCH/PUCCH transmission* on the same or different component carriers.

In principle, each physical channel is separately and independently power controlled. However, in the case of multiple physical channels to be transmitted in parallel from the same terminal, the total power to be transmitted for all physical channels may, in some cases, exceed the maximum terminal output power P_{TMAX} corresponding to the terminal power class. As will be seen below, the basic strategy is then to first ensure that transmission of any L1/L2 control signaling is assigned the power assumed to be needed for reliable transmission. The remaining available power is then assigned to the remaining physical channels.

For each uplink component carrier configured for a terminal, there is also an associated and explicitly configured *maximum per-carrier transmit power* $P_{\mathrm{CMAX},c}$, which may be different for different component carriers (indicated by the index c). Furthermore, although it does not make sense for $P_{\mathrm{CMAX},c}$ to exceed the maximum terminal output power P_{TMAX}, the sum of $P_{\mathrm{CMAX},c}$ for all configured component carriers may very well, and typically will,

Copyright Elsevier 2024

exceed P_{TMAX}. The reason is that, in many cases, the terminal will not be scheduled for uplink transmission on all its configured component carriers, and the terminal should also in that case be able to transmit with its maximum output power.

As will be seen in the next sections, the power control of each physical channel explicitly ensures that the total transmit power for a given component carrier does not exceed $P_{\mathrm{CMAX},c}$ for that carrier. However, the separate power-control algorithms do not ensure that the total transmit power for all component carriers to be transmitted by the terminal does not exceed the maximum terminal output power P_{TMAX}. Rather, this is ensured by a subsequent *power scaling* applied to the physical channels to be transmitted. This power scaling is carried out in such a way that any L1/L2 control signaling has higher priority, compared to data (UL-SCH) transmission.

If PUCCH is to be transmitted in the subframe, it is first assigned the power determined by its corresponding power-control algorithm, before any power is assigned to any PUSCH to be transmitted in parallel PUCCH. This ensures that L1/L2 control signaling on PUCCH is assigned the power assumed to be needed for reliable transmission before any power is assigned for data transmission.

If PUCCH is not transmitted in the subframe but L1/L2 control signaling is multiplexed on to PUSCH, the PUSCH carrying the L1/L2 control signaling is first assigned the power determined by its corresponding power-control algorithm, before any power is assigned to any other PUSCH to be transmitted in parallel. Once again, this ensures that L1/L2 control signaling is assigned the power assumed to be needed before any power is assigned for other PUSCH transmissions only carrying UL-SCH. Note that, in the case of transmission of multiple PUSCH in parallel (carrier aggregation), at most one PUSCH may include L1/L2 control signaling. Also, there cannot be PUCCH transmission and L1/L2 control signaling multiplexed on to PUSCH in the same subframe. Thus, there will never be any conflict between these rules.

If the remaining available transmit power is not sufficient to fulfill the power requirements of any remaining PUSCH to be transmitted, the powers of these remaining physical channels, which only carry UL-SCH, are scaled so that the total power for all physical channels to be transmitted does not exceed the maximum terminal output power.

Overall, the PUSCH power scaling, including the priority for PUSCH with L1/L2 control signaling, can thus be expressed as:

$$\sum_i w_c \cdot P_{\mathrm{PUSCH},c} \leq P_{\mathrm{TMAX}} - P_{\mathrm{PUCCH}} \tag{11.1}$$

where $P_{\mathrm{PUSCH},c}$ is the transmit power for PUSCH on carrier c as determined by the power-control algorithm (before power scaling but including the per-carrier limitation $P_{\mathrm{CMAX},c}$), P_{PUCCH} is the transmit power for PUCCH (which is zero if there is no PUCCH transmission in the subframe), and w_c is the power-scaling factor for PUSCH on carrier c ($w_c \leq 1$). For any PUSCH carrying L1/L2 control signaling the scaling factor w_c should be set to 1. For the remaining PUSCH, some scaling factors may be set to zero by decision of

Copyright Elsevier 2024

the terminal, in practice implying that the PUSCH, as well as the corresponding UL-SCH mapped to the PUSCH, are not transmitted. For the remaining PUSCH the scaling factors *wc* are set to the same value less than or equal to 1 to ensure that the above inequality is fulfilled. Thus, all PUSCH that are actually transmitted are power scaled by the same factor.

After this overview of some general rules for the power setting of different terminals, especially for the case of multiple physical channels transmitted in parallel from the same terminal, the power control carried out separately for each physical channel will be described in more detail.

11.5.2 Power control for PUCCH

For PUCCH, the appropriate received power is simply the power needed to achieve a desired—that is, a sufficiently low—error rate in the decoding of the L1/L2 control information transmitted on the PUCCH. However, it is then important to bear the following in mind:

- In general, decoding performance is not determined by the *received signal strength* but rather by the *received signal-to-interference-plus-noise ratio* (SINR). What is an appropriate received power thus depends on the interference level at the receiver side, an interference level that may differ between different deployments and which may also vary in time as, for example, the load of the network varies.
- As previously described, there are different PUCCH formats which are used to carry different types of uplink L1/L2 control information (hybrid-ARQ acknowledgements, scheduling requests, channel-state information, or combinations thereof). The different PUCCH formats thus carry different numbers of information bits per subframe and the information they carry may also have different error-rate requirements. The required received SINR may therefore differ between the different PUCCH formats, something that needs to be taken into account when setting the PUCCH transmit power in a given subframe.

Overall, power-control for PUCCH can be described by the following expression:

$$P_{\text{PUCCH}} = \min\left\{P_{\text{CMAX},c}, P_{0,\text{PUCCH}} + PL_{\text{DL}} + \Delta_{\text{Format}} + \delta\right\} \tag{11.2}$$

In the expression above, P_{PUCCH} is the PUCCH transmit power to use in a given subframe and PL_{DL} is the downlink path loss as estimated by the terminal. The "min $\{P_{\text{CMAX},c}, \ldots\}$" term ensures that the PUCCH transmit power as determined by the power control will not exceed the per-carrier maximum power $P_{\text{CMAX},c}$.

The parameter $P_{0,\text{PUCCH}}$ in expression (11.2) is a cell-specific parameter that is broadcast as part of the cell system information. Considering only the part $P_{0,\text{PUCCH}} + PL_{\text{DL}}$ in the PUCCH power-control expression and assuming that the (estimated) downlink path loss accurately reflects the true uplink path loss, it is obvious that $P_{0,\text{PUCCH}}$ can be seen as the *desired* or *target* received power. As discussed earlier, the required received power will

Copyright Elsevier 2024

depend on the uplink noise/interference level. From this point of view, the value of $P_{0,\text{PUCCH}}$ should take the interference level into account and thus vary in time as the interference level varies. However, in practice it is not feasible to have $P_{0,\text{PUCCH}}$ varying with the instantaneous interference level. One simple reason is that the terminal does not read the system information continuously and thus the terminal would not have access to a fully up-to-date $P_{0,\text{PUCCH}}$ value. Another reason is that the uplink path-loss estimates derived from downlink measurements will not be fully accurate, for example, due to differences between the instantaneous downlink and uplink path loss and measurement inaccuracies.

Thus, in practice, $P_{0,\text{PUCCH}}$ may reflect the average interference level, or perhaps only the relatively constant noise level. More rapid interference variations can then be taken care of by closed-loop power control (see below).

For the transmit power to reflect the typically different SINR requirements for different PUCCH formats, the PUCCH power-control expression includes the term Δ_{Format}, which adds a format-dependent power offset to the transmit power. The power offsets are defined such that a baseline PUCCH format, more exactly the format corresponding to the transmission of a single hybrid-ARQ acknowledgement (format 1 with BPSK modulation, as described in Section 11.4.1.1), has an offset equal to 0 dB, while the offsets for the remaining formats can be explicitly configured by the network. For example, PUCCH format 1 with QPSK modulation, carrying two simultaneous acknowledgements and used in the case of downlink spatial multiplexing, should have a power offset of roughly 3 dB, reflecting the fact that twice as much power is needed to communicate two acknowledgements instead of just a single acknowledgement.

Finally, it is possible for the network to directly adjust the PUCCH transmit power by providing the terminal with explicit power-control commands that adjust the term δ in the power-control expression above. These power-control commands are *accumulative*—that is, each received power-control command increases or decreases the term δ by a certain amount. The power-control commands for PUCCH can be provided to the terminal by two different means:

- As mentioned in Section 10.4, a power-control command is included in each downlink scheduling assignment—that is, the terminal receives a power-control command every time it is explicitly scheduled on the downlink. One reason for uplink PUCCH transmissions is the transmission of hybrid-ARQ acknowledgements as a response to downlink DL-SCH transmissions. Such downlink transmissions are typically associated with downlink scheduling assignments on PDCCH and the corresponding power-control commands could thus be used to adjust the PUCCH transmit power prior to the transmission of the hybrid-ARQ acknowledgements.
- Power-control commands can also be provided on a special PDCCH that simultaneously provides power-control commands to multiple terminals (PDCCH using DCI format 3/3A; see Section 10.4.7). In practice, such power-control commands are then typically transmitted on a regular basis and can be used to adjust the PUCCH transmit power, for

Copyright Elsevier 2024

example, prior to (periodic) uplink CSI reports. They can also be used in the case of semi-persistent scheduling (see Chapter 13), in which case there may be uplink transmission of both PUSCH (UL-SCH) and PUCCH (L1/L2 control) without any explicit scheduling assignments/grants.

The power-control command carried within the uplink scheduling grant consists of two bits, corresponding to the four different update steps −1, 0, +1, or +3 dB. The same is true for the power-control command carried on the special PDCCH assigned for power control when this is configured to DCI format 3A. On the other hand, when the PDCCH is configured to use DCI format 3, each power-control command consists of a single bit, corresponding to the update steps −1 and +1 dB. In the latter case, twice as many terminals can be power controlled by a single PDCCH. One reason for including the possibility for 0 dB (no change of power) as one power-control step is that a power-control command is included in *every* downlink scheduling assignment and it is desirable not to have to update the PUCCH transmit power for each assignment.

11.5.3 Power control for PUSCH

Power-control for PUSCH transmission can be described by the following expression:

$$P_{\mathrm{PUSCH},c} = \min\left\{P_{\mathrm{CMAX},c} - P_{\mathrm{PUCCH}}, P_{0,\mathrm{PUSCH}} + \alpha \cdot PL_{\mathrm{DL}} + 10 \cdot \log_{10}(M) + \Delta_{\mathrm{MCS}} + \delta\right\} \tag{11.3}$$

where M indicates the instantaneous PUSCH bandwidth measured in number of resource blocks and the term Δ_{MCS} is similar to the term Δ_{Format} in the expression for PUCCH power control—that is, it reflects the fact that different SINR is required for different modulation schemes and coding rates used for the PUSCH transmission.

The above expression is similar to the power-control expression for PUCCH transmission, with some key differences:

- The use of "$P_{\mathrm{CMAX},c} - P_{\mathrm{PUCCH}}$" reflects the fact that the transmit power available for PUSCH on a carrier is the maximum allowed per-carrier transmit power *after power has been assigned to any PUCCH transmission* on that carrier. This ensures priority of L1/L2 signaling on PUCCH over data transmission on PUSCH in the power assignment, as described in Section 11.5.1.
- The term $10 \cdot \log_{10}(M)$ reflects the fact that what is fundamentally controlled by the parameter $P_{0,\mathrm{PUSCH}}$ is the power *per resource block*. For a larger resource assignment, a correspondingly higher received power and thus a correspondingly higher transmit power is needed.[26]
- The parameter α, which can take a value smaller than or equal to 1, allows for so-called *partial path-loss compensation*, as described below.

[26]One could also have included a corresponding term in the expression for PUCCH power control. However, as the PUCCH bandwidth always corresponds to one resource block, the term would always equal zero.

Copyright Elsevier 2024

In general, the parameters $P_{0,\mathrm{PUSCH}}$, α, and Δ_{MCS} can be different for the different component carriers configured for a terminal.

In the case of PUSCH transmission, the explicit power-control commands controlling the term δ are included in the uplink scheduling grants, rather than in the downlink scheduling assignments. This makes sense as PUSCH transmissions are preceded by an uplink scheduling grant except for the case of semi-persistent scheduling. Similar to the power-control commands for PUCCH in the downlink scheduling assignment, the power-control commands for PUSCH are multi-level. Furthermore, also in the same way as for PUCCH power control, explicit power-control commands for PUSCH can be provided on the special PDCCH that simultaneously provides power-control commands to multiple terminals. These power-control commands can, for example, be used for the case of PUSCH transmission using semi-persistent scheduling.

Assuming α equal to 1, also referred to as *full path-loss compensation*, the PUSCH power-control expression becomes very similar to the corresponding expression for PUCCH. Thus, the network can select a modulation-and-coding scheme (MCS), and the power-control mechanism, including the term Δ_{MCS}, will ensure that the received SINR will match the SINR required for that modulation-and-coding scheme, *assuming that the terminal transmit power does not reach its maximum value*.

In the case of PUSCH transmission, it is also possible to "turn off" the Δ_{MCS} function by setting all Δ_{MCS} values to zero. In that case, the PUSCH received power will be matched to a certain MCS given by the selected value of $P_{0,\mathrm{PUSCH}}$.

With the parameter α less than 1, the PUSCH power control operates with so-called *partial path-loss compensation*—that is, an increased path loss is not fully compensated for by a corresponding increase in the uplink transmit power. In that case, the received power, and thus the received SINR per resource block, will vary with the path loss and, consequently, the scheduled modulation-and-coding scheme should vary accordingly. In the case of fractional path-loss compensation, the Δ_{MCS} function should be disabled. Otherwise the terminal transmit power would be further reduced when the modulation-and-coding scheme is reduced to match the partial path-loss compensation.

Figure 11.34 illustrates the differences between full path-loss compensation ($\alpha = 1$) and partial path-loss compensation ($\alpha < 1$). As can be seen, with partial path-loss compensation, the terminal transmit power increases more slowly than the increase in path loss (left-hand figure) and, consequently, the received power, and thus also the received SINR, is reduced as the path loss increases (right-hand figure). To compensate for this, the modulation-and-coding scheme—that is, the PUSCH data rate—should be reduced as the path loss increases.

The potential benefit of partial path-loss compensation is a relatively lower transmit power for terminals closer to the cell border, implying less interference to other cells. At the same time, this also leads to a reduced data rate for these terminals. It should also be noted that a similar effect can be achieved with full path-loss compensation by having the scheduled modulation-and-coding scheme depend on the estimated downlink path loss, which can be derived from the power headroom report, and rely on Δ_{MCS} to reduce the relative terminal

Copyright Elsevier 2024

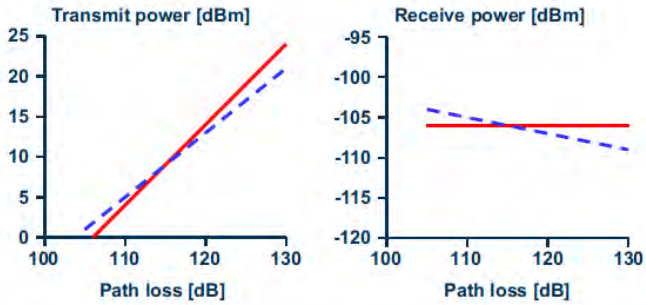


FIGURE 11.34

Full vs. partial path-loss compensation. Solid curve: full compensation ($\alpha = 1$); dashed curve: partial compensation ($\alpha = 0.8$)

transmit power for terminals with higher path loss. However, an even better approach would then be to not only base the modulation-and-coding scheme selection on the path loss to the current cell, but also on the path loss to the neighboring interfered cells.

11.5.4 Power Control for SRS

The SRS transmit power basically follows that of the PUSCH, compensating for the exact bandwidth of the SRS transmission and with an additional power offset. Thus, the power control for SRS transmission can be described according to:

$$P_{SRS} = \min\{P_{\text{CMAX},c}, P_{0,\text{PUSCH}} + \alpha \cdot PL_{\text{DL}} + 10 \cdot \log_{10}(M_{\text{SRS}}) + \delta + P_{\text{SRS}}\} \qquad (11.4)$$

where the parameters $P_{0,\text{PUSCH}}$, α, and δ are the same as for PUSCH power control, as discussed in Section 11.5.3. Furthermore, M_{SRS} is the bandwidth, expressed as number of resource blocks, of the SRS transmission and P_{SRS} is a configurable offset.

11.6 Uplink timing alignment

The LTE uplink allows for uplink intra-cell orthogonality, implying that uplink transmissions received from different terminals within a cell do not cause interference to each other. A requirement for this *uplink orthogonality* to hold is that the signals transmitted from different terminals within the same subframe but within different frequency resources (different resource blocks) arrive approximately time aligned at the base station. More specifically, any timing misalignment between received signals should fall within the cyclic prefix. To ensure such receiver-side time alignment, LTE includes a mechanism for *transmit-timing advance*.

In essence, timing advance is a negative offset, at the terminal, between the start of a received downlink subframe and a transmitted uplink subframe. By controlling the offset appropriately for each terminal, the network can control the timing of the signals received at

Copyright Elsevier 2024

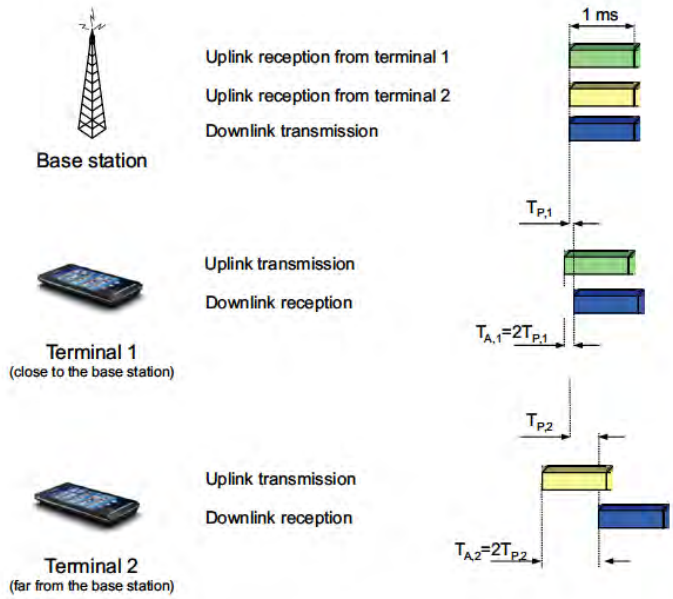


FIGURE 11.35

Uplink timing advance

the base station from the terminals. Terminals far from the base station encounter a larger propagation delay and therefore need to start their uplink transmissions somewhat in advance, compared to terminals closer to the base station, as illustrated in Figure 11.35. In this specific example, the first terminal is located close to the base station and experiences a small propagation delay, $T_{P,1}$. Thus, for this terminal, a small value of the timing advance offset $T_{A,1}$ is sufficient to compensate for the propagation delay and to ensure the correct timing at the base station. However, a larger value of the timing advance is required for the second terminal, which is located at a larger distance from the base station and thus experiences a larger propagation delay.

The timing-advance value for each terminal is determined by the network based on measurements on the respective uplink transmissions. Hence, as long as a terminal carries out uplink data transmission, this can be used by the receiving base station to estimate the uplink receive timing and thus be a source for the timing-advance commands. Sounding reference signals can be used as a regular signal to measure upon, but in principle the base station can use any signal transmitted from the terminals.

Based on the uplink measurements, the network determines the required timing correction for each terminal. If the timing of a specific terminal needs correction, the network issues a

Copyright Elsevier 2024

timing-advance command for this specific terminal, instructing it to retard or advance its timing relative to the current uplink timing. The user-specific timing-advance command is transmitted as a MAC control element (see Chapter 8 for a description of MAC control elements) on the DL-SCH. The maximum value possible for timing advance is 0.67 ms, corresponding to a terminal-to-base-station distance of slightly more than 100 km. This is also the value assumed when determining the processing time for decoding, as discussed in Section 12.1. Typically, timing-advance commands to a terminal are transmitted relatively infrequently—for example, one or a few times per second.

If the terminal has not received a timing-advance command during a (configurable) period, the terminal assumes it has lost the uplink synchronization. In this case, the terminal must re-establish uplink timing using the random-access procedure prior to any PUSCH or PUCCH transmission in the uplink.

For carrier aggregation, there may be multiple component carriers transmitted from a single terminal. The simplest way of handling this is to apply the same timing-advance value for all uplink component carriers. This is also the approach taken in release 10. In release 11, additional flexibility is provided through the introduction of so-called *timing advance groups* (TAGs) which allow different timing advance commands for different groups of component carriers. One motivation for this could be inter-band carrier aggregation, where the different component carriers are received at different geographical locations, for example by using remote radio heads for some of the bands but not others. Another example could be frequency-selective repeaters, repeating only some of the uplink component carriers.

Uplink component carriers are semi-statically grouped into timing advance groups via RRC signaling (up to four groups can be configured). All component carriers in the same group are subject to the same timing-advance command.

Copyright Elsevier 2024

Samsung Exhibit 1017, Page 323 of 538
Samsung Electronics Co., Ltd. v. Asus Technology Licensing Inc.
IPR2024-01217

This page intentionally left blank

Copyright Elsevier 202